



# **STIC Search Report**

## **Biotech-Chem Library**

STIC Database Tracking Number: 113874

TO: Manjunath N Rao  
Location: REM/3B81/3C70  
Art Unit: 1652  
February 18, 2004

Case Serial Number: 10059447

From: P. Sheppard  
Location: Remsen Building  
Phone: (571) 272-2529

sheppard@uspto.gov

### Search Notes

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 16:29:41 ; Search time 11 Seconds

(without alignments)  
123.979 Million cell updates/sec

Title: US-10-059-447a-1

Perfect score: 153

Sequence: 1 MKVLGNYIQRNFRHDKRXYTKQFNKPIX 29

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	32.7	651	1	SYM METH
2	49.5	32.4	1582	1	YU30_RALSO
3	49	32.0	521	1	CP51_CANGA
4	48	31.4	579	1	ILVB_SPLPL
5	47.5	31.0	645	1	LIP1_PHOUL
6	47	30.7	80	1	R27A_MOUSE
7	46.5	30.4	1516	1	Y819_PSBAB
8	46	30.1	271	1	NOR1_ASPPA
9	45	30.1	2514	1	PCMT_ONNVC
10	45.5	29.7	530	1	FMO3_RABIT
11	45.5	29.7	623	1	HTPG_CLOPE
12	45.5	29.7	633	1	YR45_CABEL
13	45	29.4	255	1	THID_AQUAE
14	45	29.4	256	1	TPIS_VIBPA
15	45	29.4	302	1	SC14_CANGA
16	45	29.4	350	1	DHVS_SCHPO
17	45	29.4	512	1	EMRB_ECOLI
18	45	29.4	763	1	PEPX_LACIA
19	45	29.4	1319	1	SOS1_MOUSE
20	45	29.4	1333	1	SOS1_HUMAN
21	44.5	29.1	317	1	PLC_LISMO
22	44.5	29.1	864	1	TF2_BUCAL
23	44	28.8	95	1	Y060_NPYLD
24	44	28.8	98	1	SY19_HUMAN
25	44	28.8	185	1	KITH_CFBPV
26	44	28.8	186	1	KITH_CFBPV
27	44	28.8	343	1	GUNC_CLOTF
28	44	28.8	343	1	GUNC_CLOTF
29	44	28.8	601	1	ISPG_CHLMU
30	44	28.8	772	1	LPIG_DRHOM
31	44	28.8	802	1	CSD1_ECOLI
32	44	28.8	819	1	CSD1_ECOLI
33	44	28.8	1348	1	VGR2_COTUA

## ALIGNMENTS

RESULT 1	SYM_METH	STANDARD	PRT	651 AA
AC	026587			
DT	15-0UL-1998 (Rel. 36, Created)			
DT	15-0UL-1998 (Rel. 36, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Methionyl-tRNA synthetase (EC 6.1.1.10) (Methionine--tRNA ligase)			
DE	(MeRS)			
OS	MENG OR MTH587			
OC	Methanobacterium thermoautotrophicum			
OC	Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;			
OC	Methanobacteriaceae; Methanothermobacter			
OX	NCBI_TaxID=187420;			
RP	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Delta H;			
RX	MEDLINE=98037514; PubMed=9371463;			
RA	Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,			
RA	Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,			
RA	Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,			
RA	Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,			
RA	Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,			
RA	McDougal S., Shimer G., Goyal A., Pietrovski S., Church G.M.,			
RA	Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;			
RT	"Complete genome sequence of Methanobacterium thermoautotrophicum			
RT	deltah: functional analysis and comparative genomics."			
RT	J. Bacteriol. 179:7135-7155 (1997).			
CC	-1- FUNCTION: Is required not only for elongation of protein synthesis			
CC	but also for the initiation of all mRNA translation through			
CC	initiator tRNA(Met) aminoacylation (By similarity).			
CC	-1- CATALYTIC ACTIVITY: ATP + L-methionine + tRNA(Met) = AMP +			
CC	diphosphate + L-methionyl-tRNA(Met).			
CC	-1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).			
CC	-1- SUBUNIT: Homodimer (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic.			
CC	-1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.			
CC	Meng subfamily 1.			
CC	-1- SIMILARITY: Contains 1 tRNA-binding domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as there is no way in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	-----			
DR	EMBL; AE000841; AAB85093.1; -			
DR	PIR; G69177; G69177.			
DR	HSSP; P00959; IMEA.			
DR	HAMAP; MF_00098; fused; 1.			
DR	InterPro; IPR004495; MetG_Cterm.			
DR	InterPro; IPR002300; tRNA-synt_1a.			
DR	InterPro; IPR001412; tRNA-synt_1.			
DR	InterPro; IPR002304; tRNA-synt_mec.			

DR InterPro; IPR002547; tRNA bind.  
 DR Pfam; PF00133; tRNA-synt 1; 1.  
 DR Pfam; PF01588; tRNA bind 1.  
 DR PIRSF; PIRSF001528; tRNA dimerising; 1.  
 DR PRINTS; PR01041; TRNASYNTHET.  
 DR TIGRfam; TIGR00398; metG\_1.  
 DR TIGRfam; TIGR00399; metG\_C-term; 1.  
 DR PROSITE; PS00178; AA tRNA\_LIGASE\_I; FALSE\_NEG.  
 DR PROSITE; PS50886; TRD; 1.  
 KM Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KM RNA-binding; tRNA-binding; Metal-binding; zinc; Complete proteome.  
 FT SITE 11 21  
 FT SITE 325 329  
 FT SITE 554 651  
 FT DOMAIN 143 143  
 FT METAL 143 143  
 FT METAL 146 146  
 FT METAL 156 156  
 FT METAL 159 159  
 SQ SEQUENCE 651 AA; 74631 MW; ABA4747D937DBE2C CRC64;

Query Match 32.7%; Score 50; DB 1; Length 651;  
 Best Local Similarity 47.1%; Pred. No. 7.8;  
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 3 VLGNVIGNFHYDGKFF 19  
 Db 380 VLGNFHLRTSFTRGFP 386

RESULT 2  
 ID YU30 RALSO STANDARD; PRT; 1582 AA.  
 AC OAKV02;  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein RSC3030 precursor.  
 GN RSC3030 OR RSC04727.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 CC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 CC Burkholderiaceae; Ralstonia.  
 CX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangent S.,  
 RA Atlar M., Billault A., Brottier P., Camus T.C., Cartolico L.,  
 RA Gaspin C., Lavie W., Moisan A., Robert C., Saurin W., Schlex T.,  
 RA Signier P., Thepault P., Whalen M., Winkler P., Levy N.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502(2002)  
 CC -1- SIMILARITY: BELONGS TO THE UPF0192 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL; AL646073; CADI6739.1; -  
 DR KW Hypothetical protein; Signal; Complete proteome.  
 FT SIGNAL 1 15  
 FT CHAIN 16 1582  
 SQ SEQUENCE 1582 AA; 170090 MW; 8683D148F5A3C2A CRC64;

Query Match 32.4%; Score 49.5; DB 1; Length 1582;  
 Best Local Similarity 33.3%; Pred. No. 23;  
 Matches 9; Conservative 8; Mismatches 7; Indels 3; Gaps 1;

Qy 5 GNVIGNFHYDGKFFTKCF---NKPI 28  
 Db 316 GVISENFYDSEIYNKLYAFTRPL 342

RESULT 3  
 ID CPS1 CANMA STANDARD; PRT; 521 AA.  
 AC Q12586;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cytochrome P450 52A9 (EC 1.14.14.-) (CYPLI49) (Alkane-inducible  
 DE P450-ALK5-A).  
 DE CYP52A9.  
 OS Candida maltosa (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 CX NCBI\_TaxID=5479;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95169275; PubMed=7865134;  
 RA Okuma M., Maracka S., Tanimoto T., Fujii M., Ohta A., Takagi M.;  
 RT "CYP52 (cytochrome P450alk) multigene family in Candida maltosa:  
 RT identification and characterization of eight members";  
 RL DNA Cell Biol. 14:163-173(1995).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=96311366; PubMed=8713123;  
 RA Zimmer T., Okuma M., Ohta A., Takagi M., Schunck W.H.;  
 RT "The CYP52 multigene family of Candida maltosa encodes functionally  
 RT diverse n-alkane-inducible cytochromes P450".  
 RL Biochem. Biophys. Res. Commun. 224:784-789(1996).  
 CC -1- FUNCTION: TOGETHER WITH AN NADPH CYTOCHROME P450 THE ENZYME SYSTEM  
 CC CATALYZES THE TERMINAL HYDROXYLATION AS THE FIRST STEP IN THE  
 CC ASSIMILATION OF ALKANES AND FATTY ACIDS.  
 CC -1- CATALYTIC ACTIVITY: Hydroxylation of N-alkanes at the terminal  
 CC position.  
 CC -1- INDUCTION: By N-alkanes.  
 CC -1- SIMILARITY: Belongs to the cytochrome P450 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL; D12717; BAA02211.1; -  
 DR PIR; J50723; J50723.  
 DR HSSP; P14779; IUPZ.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; p450; 1.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.  
 FT METAL 468 468  
 FT METAL 468 468  
 SQ SEQUENCE 521 AA; 59872 MW; 07D8827DC92042DA CRC64;

Query Match 32.0%; Score 49; DB 1; Length 521;  
 Best Local Similarity 40.0%; Pred. No. 8.8;  
 Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Qy 7 YIGNFHYDGKFFTKCFNK 26  
 Db 217 YDQCFDFFDGRKNFAEAFNK 236  
 ID ILVB\_SPIPL STANDARD; PRT; 579 AA.

```

AC P27868;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Acetylacetyl synthase (EC 2.2.1.6) (Acetylhydroxy-acid synthase) (ALS)
DE (fragment)
GN ILVY.
OS Spirulina platensis.
OC Bacteria; Cyanobacteria; Oscillatoriales; Spirulina.
OX NCBI_TaxID=1156;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92381487; PubMed=1512571;
RA Milano A., de Rossi E., Zanaria E., Barbierato L., Ciferri O.,
RA Ricciardi G.;
RT "Molecular characterization of the genes encoding acetylhydroxy acid
RT synthase in the cyanobacterium Spirulina platensis."
RL J. Gen. Microbiol. 138:1399-1408(1992).
CC -1- CATALYTIC ACTIVITY: 2 pyruvate = 2-acetylacetyl + CO(2).
CC -1- COFACTOR: Binds 1 thiamine pyrophosphate and 1 magnesium ion per
CC subunit (By similarity).
CC -1- PATHWAY: Valine and isoleucine biosynthesis; first step.
CC -1- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M75907; AAA26595.1; -
DR PIR; B44857; B44857.
DR HSSP; P06169; 1PVD.
DR InterPro; IPR004407; AcCoA_c19.
DR InterPro; IPR000399; Pyruvate_decarb.
DR Pfam; PF00205; TPP_enzymes; 1.
DR Pfam; PF02775; TPP_enzymes; 1.
DR Pfam; PF02776; TPP_enzymes; 1.
DR TIGRPFAM; TIGR00118; acCoA_c19; 1.
DR PROSITE; PS00187; TPP_ENZYMES; PARTIAL.
KW Transferase; Branched-chain amino acid biosynthesis; Flavoprotein;
KW Thiamine pyrophosphate.
FT ACT_SITE 61
FT NON_TER 579
SQ SEQUENCE 579 AA; 63440 MW; BC94FEA728A7889A CRC64;

Query Match 31.4%; Score 48; DB 1; Length 579;
Best Local Similarity 31.4%; Pred. No. 14;
Matches 11; Conservative 7; Mismatches 9; Indels 8; Gaps 1;

QY 2 KVLGNYIQNRYHDKKPYTK-----QFNKPI 28
DB 339 QVLGDMVLTQTYHWRKLSRNKPRNGTDLNQLREPI 373

RESULT 5
LIP1_PRODU STANDARD; PRT; 645 AA.
AC P40601;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lipase 1 precursor (EC 3.1.1.3) (Triacylglycerol lipase).
GN LIP1-1.
OS Photobacterium luminescens (Xenobacterium luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photobacterium.
OX NCBI_TaxID=29488;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-41.
RA STRAIN=KI22;

```

```

RX MEDLINE=93194791; PubMed=8449874;
RA Wang H., Downs B.C.A.;
RT "Phase variation in Xenobacterium luminescens: cloning and sequencing
RT of the lipase gene and analysis of its expression in primary and
RT secondary phases of the bacterium."
RL J. Bacteriol. 175:1665-1673(1993).
CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a
CC fatty acid anion.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE "GDSL" FAMILY OF LIPOLYTIC ENZYMES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X66379; CAA47020.1; -
DR PIR; A47081; A47081.
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR001087; Lipase_GDSL.
DR Pfam; PF03797; Autotransporter; 1.
DR Pfam; PF00657; Lipase_GDSL; 1.
DR TIGRPFAM; TIGR01414; autotrans bar1; 1.
DR PROSITE; PS01098; LIPASE_GDSL_SER; 1.
KW Hydrolase; Lipid degradation; Signal.
FT SIGNAL 1
FT CHAIN 25 645 LIPASE 1.
FT ACT_SITE 34 34 POTENTIAL.
FT ACT_SITE 330 330 POTENTIAL.
SQ SEQUENCE 645 AA; 70716 MW; 1271327BC56932F CRC64;

Query Match 31.0%; Score 47.5; DB 1; Length 645;
Best Local Similarity 38.5%; Pred. No. 18;
Matches 10; Conservative 3; Mismatches 4; Indels 9; Gaps 1;

QY 3 VLGNVIR-----NFHYDKKF 19
DB 415 LIGMISRYQDNSSPANNFYHNGY 440

RESULT 6
R27A_MOUSE STANDARD; PRT; 80 AA.
AC P49664; Q9D2W3;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 40S ribosomal protein S27a.
GN RPS27A.
OS Mus musculus (Mouse), and
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090; 10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE.
RC SPECIES=Mouse; STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96011832; PubMed=7488009;
RA Chan Y.-L., Suzuki K., Wool I.G.;
RT "Type carboxyl extensions of two rat ubiquitin fusion proteins are
RT ribosomal proteins S27a and L40."
RL Biochem. Biophys. Res. Commun. 215:682-690(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Mouse; STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,

```

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner T., Batalov S., Casavant T.,  
 RA Fleischmann W., Gasterland T., Gissi C., King B., Kocich H.,  
 RA Knehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schirml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Oikido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake U., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bull C., Fletcher C., Fujita E., Gariboldi M.,  
 RA Guncich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Togo-oka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.,  
 RT "functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).

RP SEQUENCE FROM N.A.  
 RC SPECIES=mouse;  
 RX MEDLINE=2238257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Srapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Pange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Millhys S.J.,  
 RA Bask S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huijck S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Heaton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting W., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,  
 RA Buterfield J.S.N., Krzywicki M.I., Skalska U., Smalins D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences."  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -1- MISCELLANEOUS: THIS RIBOSOMAL PROTEIN IS SYNTHESIZED AS A  
 C-TERMINAL EXTENSION PROTEIN (CEP) OF UBIQUITIN.  
 CC -1- SIMILARITY: BELONGS TO THE S27AE FAMILY OF RIBOSOMAL PROTEINS.  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X81839; CAA57432.1; ALT\_INIT.  
 DR EMBL: AK018706; BAB31357.1; ALT\_INIT.  
 DR EMBL: BC002108; AA02108.1; ALT\_INIT.  
 DR HSP: P02246; IUBI.  
 DR WGD: MG1:1925544; Rps27a.  
 DR InterPro: IPR002906; Ribosomal S27.  
 DR Pfam: PF01599; Ribosomal S27; 1.  
 KM Ribosomal protein; zinc-finger; Metal-binding.  
 FT DOMAIN 1 23 LYS-RICH (HIGHLY BASIC).  
 FT ZN FING 45 68 C4-TYPE  
 SQ SEQUENCE 80 AA; 9404 MW; F255F4B86F9DC1F CRC64;

Query Match 30.7%; Score 47; DB 1; Length 80;  
 Best Local Similarity 45.8%; Pred. No. 2.7;  
 Matches 11; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

QY 4 LGNYQRNFHYDGKFFTKQKRP 27  
 DB 56 MGSHPR--HYCGKCLITCFNKP 77

RESULT 7  
 Y919\_PSEAE STANDARD; PRT; 1516 AA.  
 AC Q9HVT2;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein PA4489 precursor.  
 GN PA4489.  
 OS Pseudomonas aeruginosa.  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 CC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltry L., Tolentino B., Westbrock-Nadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.,  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 opportunistic pathogen."  
 RT Nature 406:959-964(2000).

RL Nature 406:959-964(2000).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0192 FAMILY.  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AE004862; AA007877.1; -  
 DR PIR: F83085; F83085.  
 KM Hypothetical protein; Signal; Complete proteome.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT CHAIN 27 1516 HYPOTHEITICAL PROTEIN PA4489.  
 SQ SEQUENCE 1516 AA; 167428 MW; FDCED1561D62C18C CRC64;

Query Match 30.4%; Score 46.5; DB 1; Length 1516;  
 Best Local Similarity 29.6%; Pred. No. 60;  
 Matches 8; Conservative 9; Mismatches 7; Indels 3; Gaps 1;

QY 5 GNYQRNFHYDGKFFTKQKRP--NKPI 28  
 DB 315 GVFSNPFYDSLYDKRYIAFTDRPL 341

RESULT 8  
 NOR1\_ASPPA STANDARD; PRT; 271 AA.  
 AC Q00276;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Aflatoxin biosynthesis ketoreductase NOR-1 (EC 1.1.1.-).  
 GN NOR-1 OR NAR-1.  
 OS Aspergillus parasiticus.  
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 CC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_TaxID=5067;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 163 / NRRL 5862 / SU-1;  
 RX MEDLINE=95085270; PubMed=7993094;  
 RA Trail F., Chang P.-K., Cary J., Linz J.E.,  
 RT "Structural and functional analysis of the nor-1 gene involved in the  
 biosynthesis of aflatoxins by Aspergillus parasiticus."

```

RL Appl. Environ. Microbiol. 60:4078-4085(1994).
CC -1- FUNCTION: MAY BE INVOLVED IN THE DEHYDRATION OF NORSOLONIC ACID
CC TO FORM AVERANTIN.
CC -1- PATHWAY: Aflatoxin biosynthesis.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY. STRONG, TO E.NIDULANS SPEC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L27801; AA58798.1; -
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PROSITE; PS00061; ADH_SHORT; FALSE_NEG.
KW Oxidoreductase.
FT NP_BIND 28 52 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 185 185 BY SIMILARITY.
SQ SEQUENCE 271 AA; 29569 MW; 34059A857672059A CRC64;
Query Match 30.1%; Score 46; DB 1; Length 271;
Best Local Similarity 43.8%; Pred. No. 13;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
OY 2 KVLGNYIQNNPHYDGK 17
Db 189 KLANVYVVRKHFENK 204
RESULT 9
POINT ONVNG STANDARD; PRT; 2514 AA.
AC P1386;
DT 01-JUN-1990 (Rel. 13, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nonstructural polypeptide [Contains: Nonstructural protein NSP1;
DE Nonstructural protein NSP2; Nonstructural protein NSP3; Nonstructural
DE protein NSP4].
OS O'nyong-nyong virus (strain Gulu) (ONNV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OC NCBI_TaxId=11028;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=90177206; PubMed=2155505;
RA Levinson R.S., Strauss J.H., Strauss E.G.;
RT "Complete sequence of the genomic RNA of O'nyong-nyong virus and its
RT use in the construction of alphavirus phylogenetic trees.";
RL Virology 175:110-123(1990).
RN [2]
RP SEQUENCE OF 1334-2514 FROM N.A.
RX MEDLINE=88206074; PubMed=2834873;
RA Straus E.G., Levinson R., Rice C.M., Dalrymple J., Strauss J.H.;
RT "Nonstructural proteins nspp3 and nspp4 of Ross River and O'nyong-nyong
RT viruses: sequence and comparison with those of other alphaviruses.";
RL Virology 164:265-274(1988).
RN [3]
RP FUNCTION: NSP3 MAY BE INVOLVED IN RNA BINDING DURING REPLICATION.
RC -1- PM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M20303; AAA46784.1; -

```

```

DR PIR; A34680; MNWV2.
DR MEROPS; C09.001; -.
DR InterPro; IPR002589; ALDP.
DR InterPro; IPR002620; peptidase C9.
DR InterPro; IPR001788; RNA_dep_RNApol2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR000606; Viral_helicase1.
DR Pfam; PF01661; ALDP; 1.
DR Pfam; PF01077; peptidase C9; 1.
DR Pfam; PF00978; RNA_dep_RNApol2; 1.
DR Pfam; PF01443; Viral_helicase; 1.
DR SMART; SMO0505; ALDP; 1.
KW Polypeptide; Nonstructural protein; RNA-binding; Helicase.
FT CHAIN 1 535 NONSTRUCTURAL PROTEIN NSP1.
FT CHAIN 536 1333 NONSTRUCTURAL PROTEIN NSP2.
FT CHAIN 1334 1903 NONSTRUCTURAL PROTEIN NSP3.
FT CHAIN 1904 2514 NONSTRUCTURAL PROTEIN NSP4.
SQ SEQUENCE 2514 AA; 280114 MW; 3708374690530821 CRC64;
Query Match 30.1%; Score 46; DB 1; Length 2514;
Best Local Similarity 38.1%; Pred. No. 1.2e+02;
Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
OY 8 IQBNHYDGKCKFYTKQFNKP 28
Db 855 IVSSLHYSKRTTNEYNP 875
RESULT 10
FM03 RABIT STANDARD; PRT; 530 AA.
AC P32417;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dimethylxanthine monooxygenase [N-oxide forming] 3 (EC 1.14.13.8)
DE (Hepatic flavin-containing monooxygenase 3) (FMO 3) (Dimethylxanthine
DE oxidase 3) (FMO ID1) (FMO form 2) (FMO II).
GN FM03.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OC NCBI_TaxId=9986;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=New Zealand white; TISSUE=Liver;
RX MEDLINE=94245761; PubMed=8188717;
RA Burnett V.L., Lawton M.P., Philpott R.M.;
RT "Cloning and sequencing of flavin-containing monooxygenases FMO3 and
RT FMO4 from rabbit and characterization of FMO3.";
RL J. Biol. Chem. 269:14314-14322(1994).
RN [2]
RP SEQUENCE.
RX MEDLINE=91378577; PubMed=1898080;
RA Ozols J.;
RT "Multiple forms of liver microsomal flavin-containing monooxygenases:
RT complete covalent structure of form 2.";
RL Arch. Biochem. Biophys. 290:103-115(1991).
RN [3]
RP SEQUENCE OF 1-32.
RC TISSUE=Liver;
RX MEDLINE=89374273; PubMed=2505769;
RA Ozols J.;
RT "Liver microsomes contain two distinct NADPH-Monooxygenases with NH2-
RT terminal segments homologous to the flavin containing NADPH-
RT monooxygenase of Pseudomonas fluorescens.";
RL Biochem. Biophys. Res. Commun. 163:49-55(1989).
CC -1- FUNCTION: Involved in the oxidative metabolism of a variety of
CC xenobiotics such as drugs and pesticides. It N-oxygenates primary
CC aliphatic amines as well as secondary and tertiary amines.
CC -1- CATALYTIC ACTIVITY: N,N-dimethylxanthine + NADPH + O(2) = N,N-
CC dimethylxanthine N-oxide + NADP(+) + H(2)O.

```

```

CC -1- COFACTOR: FAD.
CC -1- SUBCELLULAR LOCATION: Microsomal.
CC -1- TISSUE SPECIFICITY: Liver.
CC -1- SIMILARITY: BELONGS TO THE FMO FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L10391; AAA21178.1; -.
CC PIR: B54096; B54096.
CC PIR: S18380; S18380.
CC InterPro: IPR001327; FAD_Pyr_redox.
CC InterPro: IPR000960; Flav_cont_mnoxgn.
CC InterPro: IPR001100; Pyr_redox.
CC Pfam: PF00743; FMO-like; 1.
CC PRINTS: PR00368; FADPR.
CC PRINTS: PR00370; FMOXYGENASE.
CC PRINTS: PR00411; PNDRDTASE1.
CC Oxidoreductase; Monooxygenase; NADP; Flavoprotein; FAD; Microsome;
CC Transmembrane; Multigene family; Polymorphism.
CC INIT_MET 0 0
CC NP_BIND 8 13 FAD (ADP PART) (POTENTIAL).
CC NP_BIND 190 195 NADP (POTENTIAL).
CC VARIANT 278 278 R -> M.
CC VARIANT 404 404 M -> V.
CC CONFLICT 75 75 D -> P (IN REF. 2).
CC CONFLICT 80 80 F -> N (IN REF. 2).
CC CONFLICT 127 129 STE -> ACC (IN REF. 2).
CC CONFLICT 171 172 HS -> RQ (IN REF. 2).
CC CONFLICT 196 196 C -> E (IN REF. 2).
CC CONFLICT 305 305 F -> EXEF (IN REF. 2).
CC CONFLICT 418 418 W -> T (IN REF. 2).
CC CONFLICT 422 422 S -> W (IN REF. 2).
CC CONFLICT 513 513 WLK -> ELM (IN REF. 2).
CC SEQUENCE 530 AA; 59684 MW; A450407620775C8 CRC64;

Query Match 29.7%; Score 45.5; DB 1; Length 530;
Best local Similarity 21.9%; Pred. No. 30;
Matches 7; Conservative 10; Mismatches 6; Indels 9; Gaps 1;

QY 5 GNYIQNF-----HYDCKFYTKQKNKP 27
DB 147 GHIVFNLPRDSPFGUKHFKGKSPHSREYKEP 178

RESULT 11
HTPG_CLOPE STANDARD; PRT; 623 AA.
AC O8XNC2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chapterone protein htpg (Heat shock protein htpg) (High temperature
DE protein G)
GN HTPG OR CPE0415.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_Taxid=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A.
RX MEDLINE=21664373; Pubmed=11792842;
RA Shimizu T., Ohtani K., Hatakeyama H., Onshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).

```

```

CC -1- FUNCTION: Molecular chaperone. Has ATPase activity (By
CC similarity).
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AP003186; BAB80122.1; -.
CC HAMAP: MF_00505; -.
CC InterPro: IPR003594; ATPbind_ATPase.
CC InterPro: IPR001404; Hsp90.
CC Pfam: PF02518; HATPase_c/1.
CC Pfam: PF00183; HSP90; 1.
CC PRINTS: PR00775; HEATSHOCK90.
CC SMART: SM00387; HATPase_c/1.
CC PROSITE: PS00298; HSP90, FALSE NEG.
CC Chapterone; ATP-binding; heat shock; Complete proteome.
CC DOMAIN 1 341 A; SUBSTRATE-BINDING (BY SIMILARITY).
CC DOMAIN 342 549 B (BY SIMILARITY).
CC DOMAIN 550 623 C.
CC SEQUENCE 623 AA; 72451 MW; 99DD4F2E8BEC652F CRC64;

Query Match 29.7%; Score 45.5; DB 1; Length 623;
Best local Similarity 45.0%; Pred. No. 35;
Matches 9; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

QY 7 YIQNFYDCKFYTKQKNKP 26
DB 284 YIPENIYD---YTTKEYEK 300

RESULT 12
YR45_CAEEL STANDARD; PRT; 633 AA.
ID YR45_CAEEL
AC Q09562;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 72.9 kDa protein P47D12.5 in chromosome III.
GN P47D12.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2.
RA Tatch A.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: WEAK, TO C.ELEGANS ZYG-11.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U22831; AAK20066.1; -.
CC PIR: P88478; P88478.
CC Wormpep; P47D12.5; CE01949.
CC Hypothetical protein.
CC SEQUENCE 633 AA; 72904 MW; A205565802F608FD CRC64;

Query Match 29.7%; Score 45.5; DB 1; Length 633;

```

Best Local Similarity 34.5%; Pred. No. 35;  
Matches 10; Conservative 6; Mismatches 10; Indels 3; Gaps 1;

Qy 3 VLGNVIOBNHY--DCKXYTKQFNKPI 28  
Db 41 INNTERRNHLYDIKSHAFSKMTXNI 69

RESULT 13  
THID\_AQUAE STANDARD; PRT; 255 AA.  
AC 067772;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Phosphomethylpyrimidine kinase (EC 2.7.4.7) (HMP-phosphate kinase) (HMP-P kinase)  
GN THID OR AQ\_1960.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=V95;  
RX MEDLINE=9816666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L., Graham D.E., Overbeek R., Shead M.A., Keller M., Aujay M., Huber R., Feldman R.A., Short J.M., Olson G.J., Swanson R.V.,  
RT "The complete genome of the hyperthermophilic bacterium Aquifex aeolicus";  
RT Nature 392:353-358(1998).  
RL Nature 392:353-358(1998).  
CC -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF HMP-P TO HMP-PP (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: ATP + 4-amino-2-methyl-5-phosphomethylpyrimidine = ADP + 4-amino-2-methyl-5-diphosphomethylpyrimidine.  
CC -1- PATHWAY: Thiamine biosynthesis.  
CC -1- SIMILARITY: BELONGS TO THE THID FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AB000765; AAC07733.1; -.  
DR PIR; H70467; H70467.  
DR InterPro; IPR004339; HMP-P\_kinase.  
DR InterPro; IPR002173; PFKB.  
DR Pfam; PF00294; PFKB; 1.  
DR TIGRfam; TIGR00097; HMP-P\_kinase; 1.  
KW Thiamine biosynthesis; Transferrase; Kinase; Complete proteome.  
SQ SQUONCE 255 AA; 28065 MW; 9544CASP662F708F CRC64;  
Query Match 29.4%; Score 45; DB 1; Length 255;  
Best Local Similarity 34.5%; Pred. No. 17;  
Matches 10; Conservative 6; Mismatches 7; Indels 6; Gaps 2;  
Qy 3 VLGNVIOBNHY--YDGKXYTKQFNKPI 27  
Db 171 IKGHLKGNVAIDILYDKSFY--EFKAP 197

GN TP1A OR VP0239.  
OS Vibrio parahaemolyticus.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; OC Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=670;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RIMD 2210633 / Serotype O3:K6;  
RX MEDLINE=22508454; PubMed=12620739;  
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagemori K., Iijima Y., Najiwa M., Nakano M., Yamashita A., Kibota Y., Kimura S., Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;  
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism distinct from that of V. cholerae";  
RT Lancet 361:743-749(2003).  
CC -1- CATALYTIC ACTIVITY: D-glyceridehyde 3-phosphate = glyceroe phosphate.  
CC -1- PATHWAY: Plays an important role in several metabolic pathways.  
CC -1- SUBUNIT: Homodimer (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AP005073; BAC58502.1; -.  
DR HAMAP; MF\_00147; -1.  
DR PROSITE; PS00171; TIM; 1.  
KW Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;  
KW Pentose shunt; Complete proteome.  
FT ACT SITE 97 97 BY SIMILARITY.  
FT ACT SITE 169 169 BY SIMILARITY.  
SQ SEQUENCE 256 AA; 26932 MW; 50A14E37B34CFA7 CRC64;  
Query Match 29.4%; Score 45; DB 1; Length 256;  
Best Local Similarity 30.4%; Pred. No. 17;  
Matches 7; Conservative 7; Mismatches 9; Indels 0; Gaps 0;  
Qy 3 VLGNVIOBNHYDCKXYTKQFN 25  
Db 94 IIGSHREKREHNSDEFIAKFN 116

RESULT 15  
SC14\_CANGA STANDARD; PRT; 302 AA.  
AC P53989;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE SC14 cytosolic factor (Phosphatidylinositol/phosphatidylcholine transfer protein) (PI/PC TP).  
GN SC14.  
OS Candida glabrata (Yeast) (Torulopsis glabrata).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; OC Saccharomycetales; mitospotic Saccharomycetales; Candida.  
OX NCBI\_TaxID=5478;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCYC 350;  
RA Dundon W., Islam K.;  
RT Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: REQUIRED FOR TRANSPORT OF SECRETORY PROTEINS FROM THE GOLGI COMPLEX. CATALYZES THE TRANSFER OF PHOSPHATIDYLINOSITOL AND PHOSPHATIDYLCHOLINE BETWEEN MEMBRANES IN VITRO (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH THE GOLGI COMPLEX AS A PERIPHERAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: Contains 1 CRAL-TRIO domain.  
CC -----



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X97320; CAA65985.1; -.  
 DR HSSP; P24280; IAU.  
 DR InterPro; IPR001251; CRAL\_TRIO.  
 DR InterPro; IPR001071; RetBind/tocTrans.  
 DR Pfam; PF00650; CRAL\_TRIO; 1.  
 DR PRINTS; PRO0180; CRETINALDHP.  
 DR SMART; SM00516; SEC14; 1.  
 DR PROSITE; PS50191; CRAL\_TRIO; 1.  
 DR Transport; Protein transport; Golgi stack.  
 KW DOMAIN 97 270  
 FT  
 SQ SEQUENCE 302 AA; 34292 MW; A857E3C70DBA2DDF CRC64;

Query Match 29.4%; Score 45; DB 1; Length 302;  
 Best Local Similarity 36.0%; Pred. No. 20;  
 Matches 9; Conservative 6; Mismatches 6; Indels 4; Gaps 1;

QY 6 NYIQRNHYDGK---XFTKQFNK 26  
 Db 87 NTIMQDFHYDEKPLVAKYPPQYVHK 111

Search completed: February 10, 2004, 16:55:08  
 Job time : 13 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 16:49:13 / Search time 35 Seconds

(Without alignments)  
213.815 Million cell updates/sec

Title: US-10-059-447A-1

Sequence: 1 MKVLGVNIQRNPHYDGKFFYTKQFNKPIX 29

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_protist:\*
- 12: sp\_virus:\*
- 13: sp\_unclassified:\*
- 14: sp\_virus:\*
- 15: sp\_bacteriophage:\*
- 16: sp\_bacteriophage:\*
- 17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	126	82.4	629	16	Q8XNK8 Clostridium
2	56	36.6	2474	12	Q8JUX6 Chikungunya
3	54	35.3	687	16	Q8ENJ1 Oceanobacill
4	53	34.6	510	3	Q8NTU7 Candida alb
5	53	34.6	963	12	Q8Y9U2 African swi
6	52.5	34.3	178	8	Q9G902 Ochomona
7	52	34.0	512	10	Q8L673 Arabidopsis
8	52	34.0	513	10	Q9CAT3 Arabidopsis
9	52	34.0	527	10	Q9JUG1 Arabidopsis
10	52	34.0	535	10	Q9LYS8 Arabidopsis
11	52	34.0	750	10	Q8VC40 Mus musculu
12	52	34.0	782	4	Q9NVT2 Homo sapien
13	52	34.0	785	11	Q8BVN5 Mus musculu
14	52	34.0	1392	4	Q9H3T8 Homo sapien
15	49.5	32.4	202	15	Q9E4J1 Human immun
16	49.5	32.4	516	5	O61758 Caenorhabdi

17	49.5	32.4	1050	5	Q9NJN5 Neospora ca
18	49.5	32.4	1051	5	Q97123 Toxoplasma
19	49.5	32.4	1051	8	Q9MTD3 Toxoplasma
20	49	32.0	42	16	Q8E183 Leptospira
21	49	32.0	355	16	Q9A8M6 Caulobacter
22	49	32.0	374	12	Q9YMP6 Lymnaea d
23	49	32.0	435	16	Q8DG00 Synchococc
24	49	32.0	436	16	Q8Z037 Arabidopsis
25	48.5	31.7	440	16	Q8U821 Agrobacteri
26	48	31.4	76	12	Q9YVY4 Arabidopsis
27	48	31.4	91	16	Q8XXK6 Arabidopsis
28	48	31.4	126	9	Q9T1F5 Arabidopsis
29	48	31.4	138	2	Q9JP27 Pseudomonas
30	48	31.4	179	2	Q8GMJ9 Escherichia
31	48	31.4	185	16	Q33117 Mycobacteri
32	48	31.4	527	5	P90932 Caenorhabdi
33	48	31.4	815	16	Q9ZKT4 Arabidopsis
34	48	31.4	1478	5	Q8IKO3 Arabidopsis
35	48	31.4	1819	12	Q8QHM4 Arabidopsis
36	48	31.4	2437	12	Q8Q273 Arabidopsis
37	48	31.4	5922	5	Q8IIN2 Arabidopsis
38	48	31.4	9271	5	Q8IES7 Arabidopsis
39	47.5	31.4	788	3	Q8XIN3 Arabidopsis
40	47	30.7	263	3	Q8NJO0 Arabidopsis
41	47	30.7	493	3	Q8TID7 Arabidopsis
42	47	30.7	962	12	Q8Y443 Arabidopsis
43	47	30.7	1532	5	Q8IC18 Arabidopsis
44	47	30.7	1591	5	Q9VRP4 Arabidopsis
45	46.5	30.4	334	2	Q8GKX3 Arabidopsis

## ALIGNMENTS

### RESULT 1

ID Q8XNK8 PRELIMINARY; PRT; 629 AA.

AC Q8XNK8;

DT 01-MAR-2002 (TREMBLrel. 20, last sequence update)

DT 01-MAR-2002 (TREMBLrel. 22, last annotation update)

DE Hypothetical protein CPE0325 (Alpha-N-acetylglucosaminidase).

GN CPE0325 OR AAAA.

OS Clostridium perfringens.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

CC Clostridium.

OK NCBI\_taxid=1502;

RN NCBI [1]

RP SEQUENCE FROM N.A.

RC STRAIN=13 / Type A;

RX PubMed=11792842;

RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,

Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayaishi H.,

"Complete genome sequence of Clostridium perfringens, an anaerobic

flesh-eater." Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 10543;

RA "Calcium M.T." Hsieh H.-Y., Chapman L.F., Smith D.S.;

"Identification, molecular cloning and expression of an alpha-N-

acetylglucosaminidase gene from Clostridium perfringens."

RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF003186; BAB80031.1; -

DR EMBL; AY121611; AAM55479.1; -

SO HYPOTHETICAL PROTEIN; Complete proteome.

SEQUENCE 629 AA; 74059 MW; A3B83FD3DD01456A CRC64;

Query Match 82.4%; Score 126; DB 16; Length 629;

Best Local Similarity 82.1%; Pred. No. 6,4e-10;

Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 MKVLGVNIQRNPHYDGKFFYTKQFNKPI 28

Db 1 MKVLGNVIGRNHFYDGKSFYTTSPINPI 28

## RESULT 2

Q8JUX6 PRELIMINARY; PRT; 2474 AA.  
 ID Q8JUX6  
 AC Q8JUX6;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Chikungunya virus.  
 OS Chikungunya virus.  
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
 OC Alphavirus.  
 CX NCBI\_TaxID=37124;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S27-African prototype;  
 RA Khan A.H., Morita K., Parquet M.C., Hasebe F., Mathenge E.G.M.,  
 RA Igatahi A.;  
 RT "Complete nucleotide sequence of Chikungunya virus and evidence of an  
 internal polyadenylation site."  
 RL J. Gen. Virol. 0:0-0(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S27-African prototype;  
 RA Khan A.H., Morita K., Parquet M.C., Hasebe F., Mathenge E.G.M.,  
 RA Igatahi A.;  
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RL EMBL, AF369024; AAN05101.1; -  
 DR InterPro; IPR002589; A1DP.  
 DR InterPro; IPR002620; Peptidase\_C9.  
 DR InterPro; IPR001788; RNA\_dep\_RNApol2.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR InterPro; IPR006066; Viral\_helicase1.  
 DR Pfam; PF01661; A1DP; 1.  
 DR Pfam; PF01707; Peptidase\_C9; 1.  
 DR Pfam; PF00978; RNA\_dep\_RNApol2; 1.  
 DR Pfam; PF01443; Viral\_helicase1; 1.  
 DR SMART; SMO0506; A1DP; 1.  
 DR PROSITE; PS50507; RDRP\_POSITIVE; 1.  
 DR PROSITE; PS50521; RDRP\_VIRAL; 1.  
 KM Polypeptid.  
 SQ SEQUENCE 2474 AA; 275649 MW; 97D250B9E5A3B0 CRC64;

Query Match 36.6%; Score 56; DB 12; Length 2474;  
 Best Local Similarity 47.6%; Pred. No. 40;  
 Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 8 IQRNHFYDGKSFYTKQFNKP 28  
 DB 855 IVSSLHIEGKMKRTTNEIKPI 875

## RESULT 3

Q8EN31 PRELIMINARY; PRT; 687 AA.  
 ID Q8EN31  
 AC Q8EN31;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Catalase (EC 1.11.1.6).  
 GN KATB OR O82659.  
 OS Oceanobacillus theysensis.  
 CC Bacteria; Firmicutes; Bacillales; Oceanobacillus.  
 CX NCBI\_TaxID=182710;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HT831 / DSM 14371 / JCM 11309;  
 RX MEDLINE=2220767; PubMed=1235376;  
 RA Takami H., Takaki Y., Uchiyama I.;

RT "Genome sequence of Oceanobacillus theysensis isolated from the Iheya  
 Ridge and its unexpected adaptive capabilities to extreme  
 environments."  
 RT Nucleic Acids Res. 30:3927-3935(2002).  
 DR EMBL; AP004602; BAC14615.1; -  
 KM Oxidoreductase; Peroxidase; Complete proteome.  
 SQ SEQUENCE 687 AA; 77756 MW; FCCB8C8267110A0 CRC64;

Query Match 35.3%; Score 54; DB 16; Length 687;  
 Best Local Similarity 46.2%; Pred. No. 20;  
 Matches 12; Conservative 4; Mismatches 8; Indels 2; Gaps 1;

QY 5 GNYIQRNHFYDGKSFYTKQFN-KPI 28  
 DB 615 GDHINDQFAYEAGNFTQFNHYKPI 640

## RESULT 4

Q8NCU7 PRELIMINARY; PRT; 510 AA.  
 ID Q8NCU7  
 AC Q8NCU7;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Gag.  
 GN Gag.  
 OS Candida albicans (Yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 CX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 10261;  
 RA Goodwin T.C.D., Foulter R.T.M.;  
 RT "Tca3 and Tcd3: unusual Ty3/gypsy-like retrotransposons from Candida  
 albicans and Candida dubliniensis."  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF510498; AAM43936.1; -  
 SQ SEQUENCE 510 AA; 59231 MW; 0BAFBBC72A72A1E CRC64;

Query Match 34.6%; Score 53; DB 3; Length 510;  
 Best Local Similarity 50.0%; Pred. No. 21;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 10 RNFHYDGKSFYTKQFNKP 27  
 DB 463 RNSNTNPKTFYPRPDKP 480

## RESULT 5

Q8V9U2 PRELIMINARY; PRT; 963 AA.  
 ID Q8V9U2  
 AC Q8V9U2;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE RNA helicase.  
 GN L09CL.  
 OS African swine fever virus (ASFV).  
 CC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfarvirus.  
 CX NCBI\_TaxID=10497;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Malawi L11-20/1;  
 RA Roberts P.C., Lu Z., Rock D.L.;  
 RT "Nucleotide sequence and analysis of 16.25 kilobase pairs of the  
 African swine fever virus genome that span the central variable  
 region."  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; L00966; AAL31320.1; -  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002464; DEAD\_Box.  
 DR InterPro; IPR001650; Helicase\_C.

DR Pfam; PF00271; helicase\_C.1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELIC\_C.1.  
 DR PROSITE; PS00690; DEAH\_ATP\_HELICASE; 1.  
 KM ATP-binding; Helicase; Hydrolase.  
 SQ SEQUENCE 963 AA; 109831 MW; C1137DAB22E2810F CRC64;

Query Match 34.6%; Score 53; DB 12; Length 963;  
 Best Local Similarity 50.0%; Pred. No. 40;  
 Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 4 LGNYIQNFHYDGKXF 19  
 DB 786 LHNIVKTHFHSKSP 801

## RESULT 6

Q9G902 PRELIMINARY; PRT; 178 AA.  
 AC Q9G902;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Ribosomal protein L6.  
 GN RPL6.  
 OS Ochromonas danica.  
 OC Mitochondrion.  
 OC Eukaryota; Stramenopiles; Chrysophyceae; Ochromonadales;  
 OC Ochromonadaceae; Ochromonas.  
 OX NCBI\_TaxID=2986;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Burger G., Lang B.F., Gray M.W.;  
 RT "Phylogenetic relationships of stramenopile algae, based on complete  
 RT mitochondrial genome sequences.";  
 RT submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AF287134; AAG18403.1;  
 DR InterPro; IPR00702; Ribosomal\_L6.  
 DR Pfam; PF00347; Ribosomal\_L6; 1.  
 DR PRINTS; PR00059; RIBOSOMALL6.  
 DR Prodom; PD002236; Ribosomal\_L6; 1.  
 KM Mitochondrion.  
 SQ SEQUENCE 178 AA; 20956 MW; 8213C3F9DF3CD9DF CRC64;

Query Match 34.3%; Score 52.5; DB 8; Length 178;  
 Best Local Similarity 38.9%; Pred. No. 8.1;  
 Matches 14; Conservative 3; Mismatches 6; Indels 13; Gaps 2;

QY 4 LGNYIQNFH-----YDGKXF-----YTKQPNK 26  
 DB 142 LGNFIQKIFHLRSYDCYKGGKFSFENKIKLKQPK 177

## RESULT 7

Q8LG73 PRELIMINARY; PRT; 512 AA.  
 AC Q8LG73;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical protein.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Haas B.J., Volkovskiy N., Town C.D., Troukhan M., Alexandrov N.,  
 RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;  
 RT "Full-length messenger RNA sequences greatly improve genome  
 RT annotation.";  
 RL Genome Biol. 0:0-0(2002).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
 RA Feldmann K.;  
 RT "Full-length cDNA from Arabidopsis thaliana.";  
 RT submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY084430; AAM61003.1;  
 KM Hypothetical protein.  
 SQ SEQUENCE 512 AA; 56168 MW; 34EA672AE3C151F5 CRC64;

Query Match 34.0%; Score 52; DB 10; Length 512;  
 Best Local Similarity 53.3%; Pred. No. 29;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 12 FHYDGKRYTKQFNK 26  
 DB 388 FNYDGKRYTKQFPER 402

## RESULT 8

Q9CAT3 PRELIMINARY; PRT; 513 AA.  
 AC Q9CAT3;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical 56.3 kDa protein.  
 GN T18K17.15 OR AT1G73180.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
 RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Egu P., Feldhym T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gill U.S., Goldsmith A.D., Haas B., Hansen M.F., Hughes B., Hultzer L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 RA Kim C.J., Koo H.L., Kremetskaia I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,  
 RA Miltseher J., Miranda M., Nguyen M., Nielsen W.C., Osborne B.I.,  
 RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Utechtack T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 408:816-820(2000).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banb J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,  
 RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,  
 RA Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,  
 RA Ishida J., Jones T., Kamita A., Karlin-Neumann G., Kawai J., Kim C.,  
 RA Koesema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,  
 RA Nguyen M., Palm C.J., Sakurai T., Saito M., Seki M., Shinn P.,  
 RA Southwick A., Tracy S.E., Shinzaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RT "Full length cDNA of gene T18K17.15 (GI:12324325).";  
 RT submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banb J., Chan M.M., Chang C.H., Chang E., Dale J.M.,  
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,  
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,

RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamita A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Nguyen M.,  
 RA Sakurai T., Satou M., Seki M., Shim P., Southwick A., Shinzaki K.,  
 RA Davis R.W., Ecker J.R., Theologis A., Clones."  
 RT "Arabidopsis Open Reading Frame (ORF) Clones."  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC010556; AAC52134.1; -  
 DR EMBL; AY056302; AL07151.1; -  
 DR EMBL; AY091307; AAM14246.1; -  
 KM Hypothetical protein.  
 SQ SEQUENCE 513 AA; 56297 MW; 21B221CE67DE8DAD CRC64;

Query Match 34.0%; Score 52; DB 10; Length 513;  
 Best Local Similarity 53.8%; Pred. No. 29;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 12 FHVGGKXFTKQFNK 26  
 Db 389 FNDGKRYKKMFER 403

RESULT 9  
 Q9SGJ1 PRELIMINARY; PRT; 527 AA.

AC Q9SGJ1 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Putative chloroplast nucleoid DNA binding protein.  
 GN ATG342980.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids 1; Brassicales; Brassicaceae; Arabidopsis.  
 NC NCB1\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ncv. Columbia;  
 RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., Vanden S.E.,  
 RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,  
 RA Carreira A.J., Creasy T.H., Buell C.R., Town C.D., Nerman W.C.,  
 RA Fraser C.M., Venter J.C.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC006931; AAD21712.2; -  
 DR EMBL; ACC06580; AAM15292.1; -  
 DR HSSP; P00799; 2ASI.  
 DR InterPro; IPR001461; AspproteaseA1.  
 DR InterPro; IPR001969; Aspprotease\_site.  
 DR Pfam; PF00026; asp; 1.  
 DR PRINTS; PR00792; PEPsin.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 527 AA; 58649 MW; B662162BBA6E4BD CRC64;

Query Match 34.0%; Score 52; DB 10; Length 527;  
 Best Local Similarity 58.8%; Pred. No. 30;  
 Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

Qy 3 VLGNYIQRNFR-YDGK 17  
 Db 497 IIGNYQQNFHLYDTK 513

RESULT 10  
 Q9LYS8 PRELIMINARY; PRT; 535 AA.

AC Q9LYS8 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Hypothetical protein (AT3G59080/F17J16\_130).  
 GN F17J16\_130 OR AT3G59080.  
 OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids 1; Brassicales; Brassicaceae; Arabidopsis.  
 NC NCB1\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,  
 RA Rudd S., Lemcke K., Mayer K.F.X., Queller F., Salanoubat M.;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Shin P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,  
 RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,  
 RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamita A.,  
 RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,  
 RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,  
 RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,  
 RA Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamita A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinzaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Nguyen M., Karlin-Neumann G., Southwick A., Timp M., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamita A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinzaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL163527; CAB86936.1; -  
 DR EMBL; AF424562; AAL11556.1; -  
 DR EMBL; AY099818; AAM20659.1; -  
 DR EMBL; BT000326; AAN15645.1; -  
 DR HSSP; P00799; 2ASI.  
 DR InterPro; IPR001461; AspproteaseA1.  
 DR InterPro; IPR001969; Aspprotease\_site.  
 DR Pfam; PF00026; asp; 1.  
 DR PRINTS; PR00792; PEPsin.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 535 AA; 59242 MW; 4C20AF38A2B66472 CRC64;

Query Match 34.0%; Score 52; DB 10; Length 535;  
 Best Local Similarity 58.8%; Pred. No. 30;  
 Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

Qy 3 VLGNYIQRNFR-YDGK 17  
 Db 505 IIGNYQQNFHLYDTK 521

RESULT 11  
 Q8VC40 PRELIMINARY; PRT; 750 AA.

AC Q8VC40 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Hypothetical 82.9 Kda protein.

```

GN BC021875.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL: BC021875; AAH21875.1; -.
DR MGD; MGI:2384296; BC021875.
DR InterPro; IPR005054; RNA_rec_mot.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Hypothetical protein.
SQ SEQUENCE 750 AA; 82874 MW; 7335ECFDFC2460 CRC64;

Query Match
Best Local Similarity 34.0%; Score 52; DB 11; Length 750;
Matches 12; Conservative 4; Mismatches 12; Indels 10; Gaps 1;

QY 1 MKVLGNVYIQNPFHYDGKXF-----YTKQFNKPI 28
DB 488 MKLRGMVYIARQLSFTGVTFKIEEVLSSQSYVMKNKAV 525

RESULT 12
Q9NVE2 ID Q9NVE2 PRELIMINARY; PRT; 792 AA.
AC Q9NVE2;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ10833 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuma S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto T., Wakamatsu A.,
RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK001695; BA91842.1; -.
DR InterPro; IPR005054; RNA_rec_mot.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Hypothetical protein.
FT NON TER 782
SQ SEQUENCE 782 AA; 86258 MW; DOBBD5B8B8E4DF6 CRC64;

Query Match
Best Local Similarity 34.0%; Score 52; DB 4; Length 782;
Matches 12; Conservative 4; Mismatches 12; Indels 10; Gaps 1;

QY 1 MKVLGNVYIQNPFHYDGKXF-----YTKQFNKPI 28
DB 524 MKLRGMVYIARQLSFTGVTFKIEEVLSSQSYVMKNKAV 561

RESULT 13
Q8BVN5 ID Q8BVN5 PRELIMINARY; PRT; 785 AA.
AC Q8BVN5;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE CDNA FLJ10833 FTS (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK077133; BAC3636.1; -.
FT NON TER 785
SQ SEQUENCE 785 AA; 86515 MW; C8C7D22310A418C9 CRC64;

Query Match
Best Local Similarity 34.0%; Score 52; DB 11; Length 785;
Matches 12; Conservative 4; Mismatches 12; Indels 10; Gaps 1;

QY 1 MKVLGNVYIQNPFHYDGKXF-----YTKQFNKPI 28
DB 524 MKLRGMVYIARQLSFTGVTFKIEEVLSSQSYVMKNKAV 561

RESULT 14
Q9H3T8 ID Q9H3T8 PRELIMINARY; PRT; 1392 AA.
AC Q9H3T8;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE MOP-3.
GN MOP-3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Blood;
RA Takayama K., Ukai Y., Fujii Y., Yoshimoto M.;
RT "Molecular and Biological Characterization of a New Nuclear Protein,
RT MOP-3 which is highly expressed in Human Monocytes.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB014772; BAB19784.1; -.
DR InterPro; IPR005054; RNA_rec_mot.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 1392 AA; 154157 MW; F1088016981F09D1 CRC64;

Query Match
Best Local Similarity 34.0%; Score 52; DB 4; Length 1392;
Matches 12; Conservative 4; Mismatches 12; Indels 10; Gaps 1;

QY 1 MKVLGNVYIQNPFHYDGKXF-----YTKQFNKPI 28
DB 524 MKLRGMVYIARQLSFTGVTFKIEEVLSSQSYVMKNKAV 561

RESULT 15
Q9E4U1 ID Q9E4U1 PRELIMINARY; PRT; 202 AA.
AC Q9E4U1;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RA Leroux C.;
RT "HIV-1 shedding.";

```

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF256394; AAC16055.1; -.  
DR InterPro: IPR000777; GP120.  
DR Pfam: PF00516; GP120; 1.  
KW AIDS; Coat protein; Glycoprotein.  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 202 AA; 22825 MW; 89BF5A4D87F474BB CRC64;  
  
Query Match 32.4%; Score 49.5; DB 15; Length 202;  
Best Local Similarity 47.6%; Pred. No. 25;  
Matches 10; Conservative 3; Mismatches 5; Indels 3; Gaps 1;  
  
QY 8 IQRNHYDGKKFY--TKQFN 25  
: ||: ||: ||: ||: ||: ||  
Db 104 VMHNPYGGEFFYCDTTLKFN 124  
  
Search completed: February 10, 2004, 16:55:57  
Job time : 39 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 16:28:11 / Search time 41 Seconds

(without alignments)  
112.270 Million cell updates/sec

Title: US-10-059-447A-1

Perfect score: 153  
Sequence: 1 MKVLGNYIQNFHYDGKXFKTKQFKPKIX 29

Scoring table:  
BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

A Geneseq 19Jun03:\*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	34.0	459	21	AAAG20638 Arabidopsis thaliana
2	52	34.0	461	21	AAAG20637 Arabidopsis thaliana
3	52	34.0	512	21	AAAG20636 Arabidopsis thaliana
4	52	34.0	705	20	AAAY29513 Human lung tumour
5	52	34.0	705	21	AAAB44452 Human lung tumour
6	52	34.0	705	22	AAAB3803 Human lung tumour
7	52	34.0	782	22	AAAB3229 Human protein sequ
8	52	34.0	910	23	ABPE8943 Human polypeptide
9	50	32.7	765	24	ABP96229 Human nucleic-acid

10	49.5	32.4	307	23	ABBS5253
11	49	32.0	567	23	ABP73883
12	48	31.4	110	22	AAO00958
13	48	31.4	814	19	AAV11067
14	48	31.4	814	20	AAV17162
15	48	31.4	815	19	AAV11066
16	48	31.4	815	20	AAV17161
17	47	30.7	163	23	ABP51425
18	47	30.7	1500	22	ABBS5870
19	46.5	30.4	1312	18	AAW22775
20	46.5	30.4	1312	19	AAW71295
21	46	30.1	541	20	AAW89416
22	46	30.1	1307	22	ABG06086
23	46	30.1	1307	22	ABG09666
24	46	30.1	1307	22	ABG10826
25	46	30.1	1307	22	ABG17893
26	46	30.1	1356	23	AAU79430
27	46	30.1	1517	22	ABG23672
28	46	30.1	1628	22	ABG02195
29	45.5	29.7	147	22	ABBS6406
30	45.5	29.7	147	22	AAW95875
31	45	29.4	92	22	AAO06285
32	45	29.4	156	13	AAK26815
33	45	29.4	163	23	ABP42349
34	45	29.4	308	23	ABBS89328
35	45	29.4	318	23	ABP64224
36	45	29.4	372	23	ABBS9781
37	45	29.4	372	24	AAO19621
38	45	29.4	422	22	AAU15889
39	45	29.4	422	22	AAU16344
40	45	29.4	422	24	ABU54958
41	45	29.4	422	24	ABU55413
42	45	29.4	440	20	AAV17179
43	45	29.4	463	18	AAW21014
44	45	29.4	733	23	ABBS5403
45	45	29.4	1119	14	AAK47043

#### ALIGNMENTS

RESULT 1	AAAG20638	standard; Protein; 459 AA.
ID	AAAG20638;	
AC	AAAG20638;	
XX		
XX	17-OCT-2000	(first entry)
DT		
XX		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 22907.	
KW	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
XX		
OS	Arabidopsis thaliana.	
PN	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
XX		
PF	25-FEB-2000; 2000EP-0301439.	
XX		
PR	05-FEB-1999; 99US-0121825.	
PR	05-MAR-1999; 99US-0123180.	
PR	09-MAR-1999; 99US-0123548.	
PR	23-MAR-1999; 99US-0125788.	
PR	23-MAR-1999; 99US-0126264.	
PR	29-MAR-1999; 99US-0126785.	
PR	01-APR-1999; 99US-0127462.	
PR	06-APR-1999; 99US-0128234.	
PR	08-APR-1999; 99US-0128714.	
PR	16-APR-1999; 99US-0129845.	

Lactococcus lactis  
Candida albicans e  
Human polypeptide  
H. pylori ORF 13ae  
H. pylori outer me  
H. pylori ORF 13ae  
H. pylori outer me  
Human MDRT SEQ ID  
Drosophila melanog  
Human RAD50. Homo  
Moraxella catarrha  
Novel human diagno  
Novel human diagno  
Novel human diagno  
Novel human diagno  
Novel human diagno  
Novel human diagno  
Novel human diagno  
Human testicular a  
Human reproductive  
Human polypeptide  
Human ubiquitin.  
Human ovarian anti  
Human polypeptide  
Human ORF594. Hom  
Novel human protei  
Human nucleic acid  
Human novel secret  
Human novel polype  
Human novel polype  
H. pylori outer me  
H. pylori cytoplasm  
Lactococcus lactis  
Mammalian son of s



PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 26-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132853.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140685.  
PR 28-JUN-1999; 99US-0140683.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142305.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144634.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 27-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145921.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 15-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159283.  
PR 13-OCT-1999; 99US-0159284.  
PR 13-OCT-1999; 99US-0159285.  
PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159337.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160982.  
PR 22-OCT-1999; 99US-0160983.  
PR 22-OCT-1999; 99US-0160984.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 34.0%; Score 52; DB 21; Length 459;  
Best Local Similarity 53.3%; Pred. No. 23;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 12 FHYDGKXFFYTKQFNK 26

DB 335 FNYDGKRYKKKFER 349

## RESULT 2

AA020637

ID AA020637 standard; Protein; 461 AA.

XX AA020637;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 22906.

XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121925.

PR 05-MAR-1999; 99US-0123160.

PR 05-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132068.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0142977.  
PR 14-JUL-1999; 99US-0143542.  
PR 15-JUL-1999; 99US-0143624.  
PR 16-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144353.  
PR 20-JUL-1999; 99US-0144637.  
PR 20-JUL-1999; 99US-0144868.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.

```
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145216.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147312.
PR 05-AUG-1999; 99US-0147320.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 34.0%; Score 52; DB 21; Length 461;
Best Local Similarity 53.3%; Pred. No. 23;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 12 FHYDGKRYTKQFNK 26
DB 337 FNYDGKRYTKQFNK 351

RESULT 3
AAG20636
ID AAG20636 standard; Protein; 512 AA.
AC AAG20636;
XX 17-OCT-2000 (first entry)
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 22905.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130892.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134258.
```

PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134944.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137501.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140921.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0143542.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 12-AUG-1999; 99US-0148319.  
PR 13-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149368.  
PR 18-AUG-1999; 99US-0149175.  
PR 20-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149723.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150565.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-SEP-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0156369.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158223.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.



XX DE Human lung tumour-specific protein SAL-33.  
 XX KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
 XX KW antisense-therapy; vaccine; immune response; lung cancer; SAL-33.  
 OS Homo sapiens.  
 XX MO200172295-A2.  
 XX  
 XX PD 04-OCT-2001.  
 XX PF 28-MAR-2001; 2001WO-US09991.  
 XX  
 XX PR 29-MAR-2000; 2000US-0538037.  
 XX PR 05-JUN-2000; 2000US-0588937.  
 XX PR 18-AUG-2000; 2000US-0640878.  
 XX PR 22-SEP-2000; 2000US-234517P.  
 XX PR 01-NOV-2000; 2000US-0704512.  
 XX PR 14-DEC-2000; 2000US-0738973.  
 XX  
 XX PA (CORI-) CORIXA CORP.  
 XX PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indrias CY;  
 XX PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
 XX DR WPI: 2001-639201/73.  
 XX DR N-PSDB; AMD23221.  
 XX  
 XX PT New human lung-specific polynucleotides and polypeptides for the  
 XX PT diagnosis and treatment of disease e.g. lung cancer -  
 XX PS Example 4; Page 218-220; 378pp; English.  
 XX  
 XX CC The invention relates to isolated lung tumour-specific proteins and  
 XX CC their corresponding cDNA molecules. Lung tumour-specific proteins and  
 XX CC their antigen-presenting cells are useful for stimulating and/or  
 XX CC expanding T cells specific for a tumour protein, and for inhibiting  
 XX CC the development of cancer. The invention also relates to a composition  
 XX CC useful for stimulating an immune response, and for treating cancer. The  
 XX CC lung tumour specific oligonucleotide is useful in gene therapy and for  
 XX CC diagnosis, detection and treatment of lung cancer. The present sequence  
 XX CC is human lung tumour-specific protein.  
 XX  
 XX SQ Sequence 705 AA;  
 XX  
 XX Query Match 34.0%; Score 52; DB 22; Length 705;  
 XX Best Local Similarity 31.6%; Pred. No. 36;  
 XX Matches 12; Conservative 4; Mismatches 12; Indels 10; Gaps 1;  
 XX  
 XX QY 1 MKVTGNYIQNPFHYDGKX-----YTKQFNKPI 28  
 XX 451 MKLRGMYIARQLSFTGVTFKIEEVLSSQSYKMYNKAV 488  
 XX DB  
 XX  
 XX RESULT 7  
 XX ID AAB93229 standard; Protein; 782 AA.  
 XX AC AAB93229;  
 XX XX  
 XX DT 26-JUN-2001 (first entry)  
 XX DE Human protein sequence SEQ ID NO:12223.  
 XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX OS Homo sapiens.  
 XX OS EPI074617-A2.  
 XX PN 07-FEB-2001.  
 XX PD  
 XX KW

PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 XX PR 29-JUL-1999; 99JP-0248036.  
 XX PR 27-AUG-1999; 99JP-0300253.  
 XX PR 11-JAN-2000; 2000JP-0118776.  
 XX PR 02-MAY-2000; 2000JP-0183767.  
 XX PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 XX PA (HELI-) HELIX RES INST.  
 XX PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX DR WPI: 2001-318749/34.  
 XX  
 XX PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 XX PT full-length cDNAs defined in the specification, and for the detection  
 XX PT and/or diagnosis of the abnormality of the proteins encoded by the  
 XX PT full-length cDNAs -  
 XX PS Claim 8; SEQ ID 12223; 2537pp + CD ROM; English.  
 XX  
 XX CC The present invention describes primer sets for synthesizing 5602  
 XX CC full-length cDNAs defined in the specification. Where a primer set  
 XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 XX CC to the complementary strand of a polynucleotide which comprises one of  
 XX CC the 5602 nucleotide sequences defined in the specification, where the  
 XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 XX CC of an oligonucleotide comprising a sequence complementary to the  
 XX CC complementary strand of a polynucleotide which comprises a 5'-end  
 XX CC sequence and an oligonucleotide comprising a sequence complementary to a  
 XX CC polynucleotide which comprises a 3'-end sequence, where the  
 XX CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 XX CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 XX CC the specification. The primer sets can be used in antisense therapy and  
 XX CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 XX CC particularly full-length cDNAs. The primers are also useful for the  
 XX CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 XX CC the full-length cDNAs. The primers allow obtaining of the full-length  
 XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 XX CC AAH1633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 XX CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 XX CC represent oligonucleotides, all of which are used in the exemplification  
 XX CC of the present invention.  
 XX  
 XX SQ Sequence 782 AA;  
 XX  
 XX Query Match 34.0%; Score 52; DB 22; Length 782;  
 XX Best Local Similarity 31.6%; Pred. No. 41;  
 XX Matches 12; Conservative 4; Mismatches 12; Indels 10; Gaps 1;  
 XX  
 XX QY 1 MKVLGNYIQNPFHYDGKX-----YTKQFNKPI 28  
 XX 524 MKLRGMYIARQLSFTGVTFKIEEVLSSQSYKMYNKAV 561  
 XX DB  
 XX  
 XX RESULT 8  
 XX ID AAB68943 standard; Protein; 910 AA.  
 XX AC AAB68943;  
 XX XX  
 XX DT 20-JAN-2003 (first entry)  
 XX DE Human polypeptide SEQ ID NO 990.  
 XX KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
 XX KW cell-proliferative disorder; neurodegenerative diseases; bacterial;  
 XX KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
 XX KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
 XX KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;  
 XX KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
 XX KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;

KW antiarthritic.  
XX  
OS Homo sapiens.  
XX WO200270539-A2.  
FN  
XX 12-SEP-2002.  
PD  
XX 05-MAR-2002; 2002WO-US05095.  
PF  
XX 05-MAR-2001; 2001US-0799451.  
PR  
XX (HYSE-) HYSEQ INC.  
PA  
XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
PI Wehrman T, Wang J, Wang D, Drmanac RT;  
XX WPI; 2002-759812/82.  
DR N-PSDB; ABZ11160.  
DR  
XX New polynucleotides comprising sequences assembled from expressed  
PT sequence tags (ESTs), useful for treating cell-proliferative,  
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or  
PT platelet or coagulation disorders -  
XX  
PS Claim 9; SEQ ID NO 990; 1012pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated polynucleotide (I) comprising a  
CC nucleotide sequence selected from any of 948 sequences  
CC (ABZ1119-ABZ12066) or their mature protein coding portion, active domain  
CC coding protein or complementary sequences. The polynucleotides are useful  
CC for identifying expressed genes or for physical mapping of human genome.  
CC The encoded polypeptides (ABP68902-ABP69849) are useful as molecular  
CC weight markers, as a food supplement, for generating antibodies, in  
CC medical imaging, screening and diagnostic assays and for treating  
CC cell-proliferative disorders (cancer), neurodegenerative diseases  
CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple  
CC sclerosis, diabetes, lupus genetic disorders, myeloid or lymphoid  
CC disorders, platelet or coagulation disorders, wound, burns, incision,  
CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,  
CC parasitic), arthritis, etc.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 910 AA;  
QY  
Query Match 34.0%; Score 52; DB 23; Length 910;  
Best Local Similarity 31.6%; Pred. No. 48;  
Matches 12; Conservative 4; Mismatches 12; Indels 10; Gaps 1;  
DB 525 MKRGMTIAROLSTGVTFKIEVVLSCSYKMNKAV 562  
QY 1 MKVLGNTIQRNFHDKGF-----YTKQNKRI 28  
DB 525 MKRGMTIAROLSTGVTFKIEVVLSCSYKMNKAV 562  
RESULT 9  
ABP62229  
ID ABBP6229 standard; Protein; 765 AA.  
XX  
AC ABBP6229;  
XX  
DT 19-MAY-2003 (first entry)  
XX  
DE Human nucleic-acid associated protein 12 SEQ ID NO:12.  
XX  
XX Human; nucleic-acid associated protein; MAP; cardiac; cytosolic;  
KM neuroprotective; gene therapy; cardiovascular disorder; cancer;  
KM neurological disorder.  
XX  
OS Homo sapiens.  
XX

PN WO2003016549-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002WO-US25829.  
XX  
PR 17-AUG-2001; 2001US-313111P.  
PR 24-AUG-2001; 2001US-314682P.  
PR 24-AUG-2001; 2001US-314756P.  
PR 27-AUG-2001; 2001US-315105P.  
PR 31-AUG-2001; 2001US-316751P.  
PR 31-AUG-2001; 2001US-316856P.  
PR 05-OCT-2001; 2001US-328185P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
XX Barroso I, Baughn MR, Becha SD, Blake JJ, Borowsky ML, Burford N;  
PI Dugan BM, Elliott VS, Emerling BM, Foreythe ID, Gietzen KJ,  
PI Gorvad AE, Griffin JA, Hatalia AJA, Honchell CD, Ison CH, Khan FA;  
PI Lal PG, Lee EA, Lee SY, Li JX, Lu DM, Lu Y;  
PI Lehar-Mason PM, Nguyen DB, Ramkumar J, Sprague WM, Tang YT;  
PI Thangavelu K, Thornton M, Tran UK, Wallia NK, Warren BA, Xu Y;  
PI Yao MG, Yue H, Yue H, Zebardjian Y;  
DR WPI; 2003-248159/25.  
DR N-PSDB; ABZ79884.  
DR  
XX New human nucleic acid-associated proteins polypeptide, useful for  
PT preparing a composition for diagnosing or treating e.g., cardiovascular  
PT or neurological disorders -  
XX  
PS Claim 1; Page 216-217; 290pp; English.  
XX  
CC ABZ79873 to ABZ79905 encode the human nucleic-acid associated proteins  
CC designated MAP-1 to MAP-33 given in ABBP6218 to ABBP6250. The MAP  
CC sequences have cardiac, cytosolic and neuroprotective activities, and  
CC can be used in gene therapy. The MAP sequences can be used for  
CC preparing a composition for diagnosing or treating a disease or  
CC condition associated with decreased expression or overexpression of  
CC functional MAP e.g. cardiovascular or neurological disorders or cancer.  
XX  
SQ Sequence 765 AA;  
QY  
Query Match 32.7%; Score 50; DB 24; Length 765;  
Best Local Similarity 44.4%; Pred. No. 79;  
Matches 8; Conservative 6; Mismatches 4; Indels 0; Gaps 0;  
DB 209 QPNYYTKLRFYSKQYHK 226  
QY 9 QNFRHYDGKRPYTKQFNK 26  
DB 209 QPNYYTKLRFYSKQYHK 226  
RESULT 10  
ABBS5253  
ID ABB55253 standard; Protein; 307 AA.  
XX  
AC ABB55253;  
XX  
DT 16-MAY-2002 (first entry)  
XX  
DE Lactococcus lactis protein cfa.  
XX  
XX Lactococcus lactis IL1403.  
OS  
XX Lactococcus lactis IL1403.  
PN  
XX FR2807446-A1.  
PD 12-OCT-2001.  
XX  
PF 11-APR-2000; 2000FR-0004630.  
XX  
XX 11-APR-2000; 2000FR-0004630.  
XX

Query Match	32.4%;	Score 49.5;	DB 23;	Length 307;
Best Local Similarity	41.4%;	Pred. No. 35;		
Matches 12; Conservative	5;	Mismatches 5;	Indels 7;	Gaps 2

```

RESULT 11
ABP73883
ID      ABP73883  standard; Protein; 567 AA

```

KW Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis  
KW signal transduction; DNA replication; cell division; growth;  
KW proliferation; *Candida albicans*; fungicide; antifungal.

PF 26-DEC-2001; 2001WO-US49486.

PR 29-DEC-2000; 2000US-259128P.  
PR 20-FEB-2001; 2001US-0792024.  
PR 22-AUG-2001; 2001US-314050P.

PA (ELIT-) ELITRA PHARM INC.  
XX

PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;   
XX

DR WPI; 2002-566694/60.  
DR N-PSDB; ABZ32433.

PT Constructing strains for identifying gene products as effective targets  
PT for therapeutic intervention, by inactivating in the strain one allele  
PT of a gene and placing other allele of the gene under conditional  
PT expression -

The invention relates to constructing (M1) a strain of diploid fungal cells in which both alleles of a gene are modified, comprising modifying one allele by insertion or replacement by a cassette having an expressible selectable marker and modifying other allele by recombination of a promoter replacement fragment with a heterologous promoter, so that expression of the second allele is regulated by the promoter. (M1) is useful for constructing a strain of diploid fungal cells in which both alleles of a gene are modified. The diploid fungal cells having both alleles modified are useful for identifying a gene that is essential to the survival and/or pathogenicity of a fungus, a gene that contributes to the virulence and/or pathogenicity of a fungus, a gene that contributes to the resistance of a diploid fungus to an antifungal agent, an antifungal agent that inhibits the growth of a diploid fungus and for identifying a therapeutic agent for treatment of a mammalian disease. (M1) is useful for identifying a compound which modulates the activity of a gene product, preferably enzymatic activity, carbon compound catabolism, biosynthetic, transporter, transcriptional, translational, signal transduction, DNA replication and cell division activity. The method is useful for identifying a compound having the ability to inhibit growth or proliferation of *C. albicans* cells and for treating infection by *C. albicans*. The present sequence is that of an essential *Candida albicans* protein used in the method of the invention. Note: The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied to Derwent by the European Patent Office.

sq sequence 567 AA

Query Match	32.0%	Score 49;	DB 23;	Length 567;
Best Local Similarity	69.2%	Pred. No. 80;		
Matches 9;	Conservative 1;	Mismatches 3;	Indels 0;	Gaps 0

```
QY      14 YDGIKFFYTKQFNK 26
          | | | | | : | | |
Db      278 YPGKLFYTRQFIK 290
```

RESULT 12  
AAO00958  
ID AAO00958 standard; Protein; 110 AA

AC AA000958;

DT 06-NOV-2001 (first entry)

DE XX	Human polypeptide SEQ ID NO 14850.
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100
101	101
102	102
103	103
104	104
105	105
106	106
107	107
108	108
109	109
110	110
111	111
112	112
113	113
114	114
115	115
116	116
117	117
118	118
119	119
120	120
121	121
122	122
123	123
124	124
125	125
126	126
127	127
128	128
129	129
130	130
131	131
132	132
133	133
134	134
135	135
136	136
137	137
138	138
139	139
140	140
141	141
142	142
143	143
144	144
145	145
146	146
147	147
148	148
149	149
150	150
151	151
152	152
153	153
154	154
155	155
156	156
157	157
158	158
159	159
160	160
161	161
162	162
163	163
164	164
165	165
166	166
167	167
168	168
169	169
170	170
171	

KM vaccine, peptide therapy; stem cell growth factor; haematopoiesis;  
 KM humant; cytokine; cell proliferation; cell differentiation; gene therapy  
 KM tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KM nervous system disorders; arthritis; inflammation.

OS Homo sapiens

PN WO200164835-A2  
 YY

PD 07-SEP-2001

PF 26-FEB-2001; 2001WO-US04927.  
YY

PR 28-FEB-2000; 2000US-0515126.  
 DP 18-MAY-2000; 2000US-0577409

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT, XX  
DR WPI; 2001-514838/56.

PT Isolated nucleic acids and polypeptides, useful for preventing





```

RESULT 15
AAV11066
ID AAV11066 standard; Protein; 815 AA.
XX
AC AAV11066;
XX
DT 08-JUN-1999 (first entry)
XX
DE H. pylori ORF 13ae10610_156411_c3_33 outer membrane protein.
XX
KW Vaccine; probe; diagnostic; ORF; cell envelope protein;
XX secreted protein; cytoplasmic protein; cellular protein.
OS Helicobacter pylori.
XX
PN W09824475-A1.
XX
PD 11-JUN-1998.
XX
PE 05-DEC-1997; 97WO-US22104.
XX
PR 14-JUL-1997; 97US-0891928.
XX 05-DEC-1996; 96US-0759625.
PR 25-MAR-1997; 97US-0823745.
XX
PA (ASTRA ) ASTRA AB.
XX
PI Alm RA, Castriotta LM, Doig PC, Kabok Z, Smith D;
XX
DR WPI; 1998-333051/29.
DR N-PSDB; AAX30595.
XX
PT New isolated Helicobacter pylori nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of infection by
PT H. pylori and other Helicobacter species
XX
PS Claims 37, 41; Page 228-230; 339pp; English.
XX
CC Recombinant or substantially pure preparations of H. pylori polypeptides
CC are disclosed, together with the nucleic acids encoding them. In all,
CC 97 ORFs are shown. The proteins are variously cell envelope proteins,
CC cytoplasmic proteins, secreted proteins or other cellular proteins.
CC Vaccines containing the nucleic acids or proteins are claimed, as are
CC probes containing at least 8 nucleotides from the nucleic acid
CC sequences. The vaccines are useful for treating or reducing the risk of
CC H. pylori infections, and the probes can be used diagnostically for
CC detecting the presence of Helicobacter in a sample. The products are
CC also of use in screening for compounds having the ability to interfere
CC with the H. pylori life cycle or to inhibit H. pylori infection.
XX
SQ Sequence 815 AA;

```

Query Match 31.4%; Score 48; DB 19; Length 815;  
 Best Local Similarity 42.9%; Pred. No. 1.7e+02;  
 Matches 9; Conservative 4; Mismatches 6; Indels 2; Gaps 1;

```

QY 3 VLGNVIOENRHYDGKGFYTKQ 23
Db 199 IMGAY--RNHHFDALVYTHQ 217

```

Search completed: February 10, 2004, 16:54:49  
 Job time : 48 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 16:51:12 ; Search time 22 Seconds

(without alignments)  
55.773 Million cell updates/sec

Title: US-10-059-447A-1

Sequence: 1 MKVLGNLYIQRNHYDGKXFKYTKQFNKPI 29

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCUTS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	98.0	29	US-09-185-476B-1	Sequence 1, Appl1
2	52	34.0	705	US-09-370-838-186	Sequence 186, App
3	47	30.7	80	US-09-732-210-353	Sequence 353, App
4	46.5	30.4	1312	US-08-592-126-148	Sequence 148, App
5	46.5	30.4	1312	US-08-687-080-51	Sequence 51, Appl
6	46.5	30.4	1312	US-09-158-595-148	Sequence 148, Appl
7	46.5	30.4	1341	US-09-252-991A-26785	Sequence 26785, A
8	46	30.1	541	US-08-867-941-14	Sequence 14, Appl
9	46	30.1	541	US-09-074-658-14	Sequence 14, Appl
10	46	30.1	1297	US-09-107-532A-4552	Sequence 4552, Ap
11	46	30.1	2439	US-09-074-658-11	Sequence 11, Appl
12	45	29.4	1319	US-08-290-731C-2	Sequence 2, Appl1
13	45	29.4	1333	US-08-290-731C-2	Sequence 2, Appl1
14	45	29.4	1336	US-09-356-952-6	Sequence 6, Appl1
15	45	29.1	1752	US-09-865-621A-2	Sequence 2, Appl1
16	44	28.8	80	US-09-252-991A-29905	Sequence 29905, A
17	44	28.8	98	US-08-375-346A-2	Sequence 2, Appl1
18	44	28.8	98	US-08-467-123B-2	Sequence 2, Appl1
19	44	28.8	98	US-08-943-136A-2	Sequence 2, Appl1
20	44	28.8	98	US-09-635-899-2	Sequence 2, Appl1
21	44	28.8	98	PCT-US95-01780-2	Sequence 2, Appl1
22	44	28.8	181	US-09-134-001C-5665	Sequence 5665, Ap
23	44	28.8	201	US-09-134-001C-5559	Sequence 5559, Ap
24	44	28.8	2431	US-07-920-281C-2	Sequence 2, Appl1
25	44	28.8	2431	US-08-466-277-2	Sequence 2, Appl1
26	43	28.1	80	US-09-732-210-344	Sequence 344, App
27	43	28.1	119	US-08-858-207A-503	Sequence 503, App

28	43	28.1	156	2	US-09-070-060-7	Sequence 7, Appl1
29	43	28.1	156	3	US-09-051-969A-3	Sequence 3, Appl1
30	43	28.1	156	3	US-09-051-969A-4	Sequence 4, Appl1
31	43	28.1	156	3	US-09-357-746-7	Sequence 7, Appl1
32	43	28.1	384	3	US-08-530-862B-4	Sequence 4, Appl1
33	43	28.1	384	4	US-08-597-313D-4	Sequence 4, Appl1
34	43	28.1	566	4	US-09-134-001C-3431	Sequence 3431, Ap
35	43	28.1	1356	1	US-08-810-116-8	Sequence 8, Appl1
36	43	28.1	1356	2	US-07-930-548A-8	Sequence 8, Appl1
37	43	28.1	1356	2	US-09-098-707A-2	Sequence 2, Appl1
38	43	28.1	1356	4	US-09-483-539-2	Sequence 4, Appl1
39	43	28.1	1367	1	US-07-813-593-4	Sequence 4, Appl1
40	43	28.1	1367	1	US-07-977-451-6	Sequence 6, Appl1
41	43	28.1	1367	1	US-07-946-507-4	Sequence 4, Appl1
42	43	28.1	1367	1	US-08-252-517-6	Sequence 6, Appl1
43	43	28.1	1367	1	US-07-906-397A-6	Sequence 6, Appl1
44	43	28.1	1367	1	US-08-601-891-6	Sequence 6, Appl1
45	43	28.1	1367	2	US-08-443-861-2	Sequence 2, Appl1

#### ALIGNMENTS

```

RESULT 1
US-09-185-476B-1
; Sequence 1, Application US/09185476B
; Patent No. 6399749
; GENERAL INFORMATION:
; APPLICANT: Smith, Dan
; TITLE OF INVENTION: ALPHA-N-ACETYLGALACTOSAMINIDASE FROM CLOSTRIDIUM PERFRINGENS
; FILE REFERENCE: 0994.00093
; CURRENT APPLICATION NUMBER: US/09185.476B
; PRIOR FILING DATE: 1998-11-03
; PRIOR APPLICATION NUMBER: 60/064,683
; PRIOR FILING DATE: 1997-11-03
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Clostridium perfringens
; FEATURE:
; NAME/KEY: Xaa
; LOCATION: (1)..(29)
; OTHER INFORMATION: Xaa is any amino acid
US-09-185-476B-1

Query Match          98.0%; Score 150; DB 4; Length 29;
Best Local Similarity 100.0%; Pred. No. 1e-16;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MKVLGNLYIQRNHYDGKXFKYTKQFNKPI 28
DB      1 MKVLGNLYIQRNHYDGKXFKYTKQFNKPI 28

RESULT 2
US-09-370-838-186
; Sequence 186, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LONG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09370.838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/265,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289

```

SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 186  
LENGTH: 705  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-370-838-186

Query Match 34.0%; Score 52; DB 4; Length 705;  
Best Local Similarity 31.6%; Pred. No. 5.8;  
Matches 12; Conservative 4; Mismatches 12; Indels 10; Gaps 1;

QY 1 MKVLGNVYQRFHYDGKXFFYTKQFNKPI 28  
DB 451 MKLNGMYIARQLSFTGVTFKLEIVLISQSYVKMYNKAV 488

RESULT 3  
US-09-732-210-353  
Sequence 353; Application US/09732210  
Patent No. 6573361  
GENERAL INFORMATION:  
APPLICANT: Bunkers, Greg J.  
APPLICANT: Liang, Jihong  
APPLICANT: Miltanck, Cindy A.  
APPLICANT: Seale, Jeffrey W.  
APPLICANT: Mu, Yonnie S.  
TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use  
FILE REFERENCE: 38-21(15036)B  
CURRENT APPLICATION NUMBER: US/09/732,210  
CURRENT FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: US 60/169,513  
PRIOR FILING DATE: 1999-12-07  
PRIOR APPLICATION NUMBER: US 60/169,340  
PRIOR FILING DATE: 1999-12-07  
NUMBER OF SEQ ID NOS: 1/53  
SEQ ID NO: 80  
LENGTH: 80  
TYPE: PRT  
ORGANISM: Rattus norvegicus  
US-09-732-210-353

Query Match 30.7%; Score 47; DB 4; Length 80;  
Best Local Similarity 45.8%; Pred. No. 3.2;  
Matches 11; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

QY 4 LGNYIQRFHYDGKXFFYTKQFNKPI 27  
DB 56 MGSHPDR-HYCKKCLTYCFNKP 77

RESULT 4  
US-08-592-126-148  
Sequence 148; Application US/08592126  
Patent No. 5821091  
GENERAL INFORMATION:  
APPLICANT: Gregory Dolganov  
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory  
NUMBER OF INVENTION: Polypeptides  
NUMBER OF SEQUENCES: 151  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Denlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,126

FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 4600-0111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 148:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1312 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Rad50.Pro-translation of SEQ ID NO:54  
US-08-592-126-148

Query Match 30.4%; Score 46.5; DB 2; Length 1312;  
Best Local Similarity 46.9%; Pred. No. 82;  
Matches 15; Conservative 1; Mismatches 7; Indels 9; Gaps 2;

QY 1 MKVLGNVYQRFHYDGKXFFYTKQ---FNKPI 28  
DB 958 MKDIENYIQ-----DKKDYKKOKETELNKVI 984

RESULT 5  
US-08-687-080-51  
Sequence 51; Application US/08687080  
Patent No. 5965427  
GENERAL INFORMATION:  
APPLICANT: Gregory Dolganov  
TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof  
NUMBER OF SEQUENCES: 175  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Denlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/687,080  
FILING DATE: 17-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/592,126  
FILING DATE: 26-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 4600-0111.30  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1312 amino acids  
TYPE: amino acid  
STRANDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: TRANS. OF RAD50 CDNA (SEQ. 54), NT.  
INDIVIDUAL ISOLATE: 389 TO 4324  
US-08-667-080-51

Query Match 30.4%; Score 46.5; DB 2; Length 1312;  
Best Local Similarity 46.9%; Pred. No. 82;  
Matches 15; Conservative 1; Mismatches 7; Indels 9; Gaps 2;

QY 1 MKVIGNYIGRNFHDGKXFTYKQ---FNKPI 28  
Db 958 MKDIENYIQ-----DGKDYKKKXKETEINKVI 984

RESULT 6  
US-09-168-595-148  
Sequence 148; Application US/09168595  
Patent No. 6555866

GENERAL INFORMATION:  
APPLICANT: Gregory Dolganov  
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory  
NUMBER OF SEQUENCES: 151  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/168,595  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/592,126  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 4600-0111  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 148:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1312 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Rad50, pro-translation of SEQ ID NO:54  
US-09-168-595-148

Query Match 30.4%; Score 46.5; DB 4; Length 1312;  
Best Local Similarity 46.9%; Pred. No. 82;  
Matches 15; Conservative 1; Mismatches 7; Indels 9; Gaps 2;

QY 1 MKVIGNYIGRNFHDGKXFTYKQ---FNKPI 28  
Db 958 MKDIENYIQ-----DGKDYKKKXKETEINKVI 984

RESULT 7  
US-09-252-991A-26785  
Sequence 26785; Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196-136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 26785  
LENGTH: 1341  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26785

Query Match 30.4%; Score 46.5; DB 4; Length 1341;  
Best Local Similarity 29.6%; Pred. No. 84;  
Matches 8; Conservative 9; Mismatches 7; Indels 3; Gaps 1;

QY 5 GNYIGRNFHDGKXFTYKQF---NKPI 28  
Db 363 GVGVSENFYDSEIYDTKIYAFTRPL 389

RESULT 8  
US-08-867-941-14  
Sequence 14; Application US/08867941  
Patent No. 5977337

GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M  
APPLICANT: Du, Run-Pan  
APPLICANT: Wang, Qitjun  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sam & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/867,941  
FILING DATE: 03-JUN-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-681 MTS:jb  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 541 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-867-941-14

Query Match 30.1%; Score 46; DB 2; Length 541;  
Best Local Similarity 50.0%; Pred. No. 37;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 MKVIGNYIGRNFHDGKXFTYKQ---FNKPI 28  
Db 958 MKDIENYIQ-----DGKDYKKKXKETEINKVI 984

QY 5 GNYTORNFHYDGKXFX 20  
|||:|||||:  
Db 393 GNYTERYQGDGKRYH 408

RESULT 9  
US-09-074-658-14  
Sequence 14, Application US/09074658  
Patent No. 6184371  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M  
APPLICANT: Run-Pan Du  
APPLICANT: Quljun Wang  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/074,658  
FILING DATE: 08-MAY-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-795  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 541 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-074-658-14

Query Match 30.1%; Score 46; DB 3; Length 541;  
Best Local Similarity 50.0%; Pred. No. 37;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 GNYTORNFHYDGKXFX 20  
|||:|||||:  
Db 393 GNYTERYQGDGKRYH 408

RESULT 10  
US-09-107-532A-4552  
Sequence 4552, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariandello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 4552:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1297 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (B) LOCATION 1...1297  
SEQUENCE DESCRIPTION: SEQ ID NO: 4552:  
US-09-107-532A-4552

Query Match 30.1%; Score 46; DB 4; Length 1297;  
Best Local Similarity 60.0%; Pred. No. 97;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 KVLGNYTORNFHYDG 16  
|||:|||||:  
Db 101 KVLGNYTORNFHYDG 115

RESULT 11  
US-09-074-658-11  
Sequence 11, Application US/09074658  
Patent No. 6184371  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M  
APPLICANT: Run-Pan Du  
APPLICANT: Quljun Wang  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/074,658  
FILING DATE: 08-MAY-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-795  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1153  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2439 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-074-658-11

Query Match 30.1%; Score 46; DB 3; Length 2439;  
Best Local Similarity 50.0%; Pred. No. 1.9e+02;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 GNYIQRNFHYDQKRY 20  
DB 2291 GNYTERYQYDQKRYH 2306

## RESULT 12

US-08-290-731C-2  
Sequence 2, Application US/08290731C  
Patent No. 5843646

GENERAL INFORMATION:  
APPLICANT: BOWTELL, David Douglas Lawrence  
TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE  
TITLE OF INVENTION: SON OF SEVENLESS (MSOS) GENE,  
TITLE OF INVENTION: AND MSOS POLYPEPTIDES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
STREET: 2100 PENNSYLVANIA AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20037

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,731C  
FILING DATE: 17-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU93/00068  
FILING DATE: 17-FEB-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PL0921/92  
FILING DATE: 17-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: KIT, Gordon  
REGISTRATION NUMBER: 30,764  
REFERENCE/DOCKET NUMBER: O-36066  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 293-7060  
TELEFAX: (202) 293-7860  
TELEX: 6491103

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1319 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-290-731C-2

Query Match 29.4%; Score 45; DB 2; Length 1319;  
Best Local Similarity 39.1%; Pred. No. 1.4e+02;

Matches 9; Conservative 6; Mismatches 4; Indels 4; Gaps 1;  
QY 1 MKVLGNYIQRNFHYDQKRYTKQ 23  
DB 142 LKLVGNVYVNRHYE---ITKQ 160

## RESULT 13

US-09-356-952-2  
Sequence 2, Application US/09356952  
Patent No. 6117663

GENERAL INFORMATION:  
APPLICANT: Borjuck-Sjodin, Ann  
APPLICANT: Margalit, S. M.  
APPLICANT: Bor-Soggi, Dafna  
APPLICANT: Cole, Philip  
APPLICANT: Kurlyan, John  
TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE  
TITLE OF INVENTION: THEREOF  
FILE REFERENCE: 600-1-228N  
CURRENT APPLICATION NUMBER: US/09/356,952  
CURRENT FILING DATE: 1999-07-19  
EARLIER APPLICATION NUMBER: 60/093,631  
EARLIER FILING DATE: 1998-07-21  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 2  
LENGTH: 1333  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-356-952-2

Query Match 29.4%; Score 45; DB 3; Length 1333;  
Best Local Similarity 39.1%; Pred. No. 1.4e+02;  
Matches 9; Conservative 6; Mismatches 4; Indels 4; Gaps 1;

QY 1 MKVLGNYIQRNFHYDQKRYTKQ 23  
DB 142 LKLVGNVYVNRHYE---ITKQ 160

## RESULT 14

US-08-290-731C-6  
Sequence 6, Application US/08290731C  
Patent No. 5843646

GENERAL INFORMATION:  
APPLICANT: BOWTELL, David Douglas Lawrence  
TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE  
TITLE OF INVENTION: SON OF SEVENLESS (MSOS) GENE,  
TITLE OF INVENTION: AND MSOS POLYPEPTIDES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
STREET: 2100 PENNSYLVANIA AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20037

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,731C  
FILING DATE: 17-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU93/00068  
FILING DATE: 17-FEB-1993

Query Match 29.4%; Score 45; DB 2; Length 1319;  
Best Local Similarity 39.1%; Pred. No. 1.4e+02;

## ATTORNEY/AGENT INFORMATION:

NAME: KIT, Gordon  
 REGISTRATION NUMBER: 30,764  
 REFERENCE/DOCKET NUMBER: Q-36066  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 293-7060  
 TELEFAX: (202) 293-7860  
 TELEX: 6491103  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1336 amino acids  
 TYPE: amino acids  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-290-731C-6

Query Match 29.4%; Score 45; DB 2; Length 1336;  
 Best Local Similarity 39.1%; Pred. No. 1.4e+02;

Matches 9; Conservative 6; Mismatches 4; Indels 4; Gaps 1;

QY 1 MKVLGNYIQRNHFHDKKFTKQ 23

Db 139 LKLVENVVRIIRHYE---ITKQ 177

## RESULT 15

US-09-865-621A-2

Sequence 2, Application US/09865621A

Patent No. 6555348

GENERAL INFORMATION:

APPLICANT: JORGENSEN, FLEMING

APPLICANT: HANSEN, OLE CAI

APPLICANT: STOUGAARD, PETER

TITLE OF INVENTION: A NEW ENZYME ISOLATED FROM A BIFIDOBACTERIUM

FILE REFERENCE: 02405.0200

CURRENT APPLICATION NUMBER: US/09/865,621A

CURRENT FILING DATE: 2001-05-29

PRIOR APPLICATION NUMBER: 60/207,154

PRIOR FILING DATE: 2000-05-26

NUMBER OF SEQ ID NOS: 41

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 1752

TYPE: PRT

ORGANISM: Bifidobacterium bifidum

US-09-865-621A-2

Query Match 29.1%; Score 44.5; DB 4; Length 1752;  
 Best Local Similarity 47.4%; Pred. No. 2.3e+02;

Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 10 RNFHYDGKXFTKQPNKPI 28

Db 921 RSFY-SRNYVYKTKGNKPI 938

Search completed: February 10, 2004, 16:56:48  
 Job time : 42 secs





```
Patent No. US20020110563A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Henderson, Robert A.
APPLICANT: Lodes, Michael J.
APPLICANT: Fling, Steven P.
APPLICANT: Mohamath, Raedon
APPLICANT: Algate, Paul A.
APPLICANT: Secrist, Heather
APPLICANT: Indrias, Carol Yoseph
APPLICANT: Benson, Darin R.
APPLICANT: Ellicot, Mark
APPLICANT: Mannion, Jane
APPLICANT: Kalos, Michael D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR LUNG CANCER
FILE REFERENCE: 210121.475C9
CURRENT APPLICATION NUMBER: US/09/738,973
CURRENT FILING DATE: 2000-12-14
NUMBER OF SEQ ID NOS: 587
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 186
LENGTH: 705
TYPE: PRT
ORGANISM: Homo sapien
US-09-738-973-186

Query Match          34.0%; Score 52; DB 10; Length 705;
Best Local Similarity 31.6%; Pred. No. 26;
Matches 12; Conservative 4; Mismatches 12; Indels 10; Gaps 1;
```

```
Qy 1 MKVLGNTYIQNPHYDGKPF-----YTKQFNKPI 28
|||:|||||:|||||:|||||:|||||:|||||:
451 MKLRGMYIARQLSFTGVTFKIEEVLTSQSYVMKNKAV 488

RESULT 3
US-09-854-133-186
Sequence 186, Application US/09854133
Publication No. US20020183499A1
GENERAL INFORMATION:
APPLICANT: Lodes, Michael J.
APPLICANT: Mohamath, Raedon
APPLICANT: Henderson, Robert A.
APPLICANT: Benson, Darin R.
APPLICANT: Secrist, Heather
TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.475C10
CURRENT APPLICATION NUMBER: US/09/854,133
CURRENT FILING DATE: 2001-05-11
NUMBER OF SEQ ID NOS: 735
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 186
LENGTH: 705
TYPE: PRT
ORGANISM: Homo sapien
US-09-854-133-186

Query Match          34.0%; Score 52; DB 10; Length 705;
Best Local Similarity 31.6%; Pred. No. 26;
Matches 12; Conservative 4; Mismatches 12; Indels 10; Gaps 1;
```

```
Qy 1 MKVLGNTYIQNPHYDGKPF-----YTKQFNKPI 28
|||:|||||:|||||:|||||:|||||:|||||:
451 MKLRGMYIARQLSFTGVTFKIEEVLTSQSYVMKNKAV 488

RESULT 4
US-10-144-649A-186
Sequence 186, Application US/10144649A
Publication No. US20030118599A1
GENERAL INFORMATION:
```

```
APPLICANT: Lodes, Michael J.
APPLICANT: Wang, Tongtong
APPLICANT: Fan, Liqun
APPLICANT: Algate, Paul A.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR LUNG CANCER
FILE REFERENCE: 210121.475C11
CURRENT APPLICATION NUMBER: US/10/144,649A
CURRENT FILING DATE: 2002-08-21
NUMBER OF SEQ ID NOS: 749
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 186
LENGTH: 705
TYPE: PRT
ORGANISM: Homo sapien
US-10-144-649A-186
```

```
Query Match          34.0%; Score 52; DB 15; Length 705;
Best Local Similarity 31.6%; Pred. No. 26;
Matches 12; Conservative 4; Mismatches 12; Indels 10; Gaps 1;
```

```
Qy 1 MKVLGNTYIQNPHYDGKPF-----YTKQFNKPI 28
|||:|||||:|||||:|||||:|||||:|||||:
451 MKLRGMYIARQLSFTGVTFKIEEVLTSQSYVMKNKAV 488
```

```
RESULT 5
US-10-310-154-446
Sequence 446, Application US/10310154
Publication No. US20030233670A1
GENERAL INFORMATION:
APPLICANT: Edgerton, Michael D
APPLICANT: Chomet, Paul S.
APPLICANT: Adams, Thomas H.
APPLICANT: Ruff, Thomas G.
APPLICANT: Agarwal, Ameeta K.
APPLICANT: Ahrens, Jeffrey E.
APPLICANT: Ball, James A.
APPLICANT: Banu, G.
APPLICANT: Bell, Erin
APPLICANT: Boddupalli, Raghava
APPLICANT: Deikman, Jill
APPLICANT: Deng, Mollan
APPLICANT: Dong, Jintzhuo
APPLICANT: Duff, Stephen M.
APPLICANT: Galligan, Meghan M.
APPLICANT: Hinchey, Brenda S.
APPLICANT: Huang, Shihshieh
APPLICANT: Johnson, G. Richard
APPLICANT: Jung, Vincent
APPLICANT: Kretzmer, Keith A.
APPLICANT: Laccetti, Lucille B.
APPLICANT: Lai, Chao-Qiang
APPLICANT: Lee, Gary
APPLICANT: Lin, Jie-Yi
APPLICANT: Liu, Jindong
APPLICANT: Lu, Bin
APPLICANT: Luethy, Michael M.
APPLICANT: Lund, Adrian
APPLICANT: Madison, Linda L.
APPLICANT: Malloy, Kathleen A.
APPLICANT: McKiel, Christine L.
APPLICANT: Miller, Philip W.
APPLICANT: Padmarathi, Manickant
APPLICANT: Parnell, Laurence D.
APPLICANT: Start, William G.
APPLICANT: Tennessee, Dan
APPLICANT: Vidya, K.R.
APPLICANT: Wang, Haiyun
APPLICANT: Xin, Zhanqun
APPLICANT: Xu, Nanfei
APPLICANT: Yang, Chunzhi
```

```

; APPLICANT: Zeng, Xiaoping
; APPLICANT: Zhang, Qiang
; APPLICANT: Zhao, Yajuan
; APPLICANT: Zhou, Li
; TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants
; FILE REFERENCE: 38-15(52796)B
; CURRENT APPLICATION NUMBER: US/10/310,154
; CURRENT FILING DATE: 2002-12-04
; PRIOR APPLICATION NUMBER: 60/337,358
; PRIOR FILING DATE: 2001-12-04
; NUMBER OF SEQ ID NOS: 736
; SEQ ID NO 446
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Zea mays
; US-10-310-154-446

Query Match          32.7%; Score 50; DB 12; Length 510;
Best Local Similarity 52.9%; Pred. No. 37;
Matches 9; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

QY 1 MKVLGNVIOQNFH-YD 15
   |||:|||||
Db 478 MSIIQNFQQNFHYVD 494

RESULT 6
US-10-369-493-18892
; Sequence 18892, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 18892
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(430)
; OTHER INFORMATION: unsure at all Xaa locations
; US-10-369-493-18892

Query Match          32.0%; Score 49; DB 12; Length 430;
Best Local Similarity 33.3%; Pred. No. 43;
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 1 MKVLGNVIOQNFH-YD 15
   |||:|||||
Db 29 MAYVANDVETTRQSDPFTLTGFNER 55

RESULT 7
US-10-032-585-7720
; Sequence 7720, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jians
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
```

```

; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7720
; LENGTH: 567
; TYPE: PRT
; ORGANISM: Candida albicans
; US-10-032-585-7720

Query Match          32.0%; Score 49; DB 12; Length 567;
Best Local Similarity 69.2%; Pred. No. 58;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 14 YDGEKFFYTKQFNK 26
   |||:|||||
Db 278 YPGKLFYTRQPIK 290

RESULT 8
US-10-369-493-5186
; Sequence 5186, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5186
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; US-10-369-493-5186

Query Match          31.4%; Score 48; DB 12; Length 527;
Best Local Similarity 33.3%; Pred. No. 75;
Matches 9; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY 2 KVLGNVIOQNFH-YD 28
   |||:|||||
Db 381 KVIDRHINRNPLNKETLFEKSDKPL 407

RESULT 9
US-10-393-602-148
; Sequence 148, Application US/10393602
; Publication No. US20030170714A1
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; Polypeptides
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESS: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICANT: US/10/393,602  
FILING DATE: 19-Mar-2003  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,126  
FILING DATE: 26-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 4600-0111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 148:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1312 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Rad50, pro-translation of SEQ ID NO:54  
SEQUENCE DESCRIPTION: SEQ ID NO: 148:  
US-10-393-602-148  
Query Match 30.4%; Score 46.5; DB 12; Length 1312;  
Best Local Similarity 46.9%; Pred. No. 3.3e+02;  
Matches 15; Conservative 1; Mismatches 7; Indels 9; Gaps 2;  
CY 1 MKVLGNYIQNFHYDGKXFYTKQ---FNKPI 28  
DB 958 MKDIENYIQ-----DKKDYKKQKXETELNKVI 964  
RESULT 10  
US-10-369-493-19944  
Sequence 19944, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 19944  
LENGTH: 584  
TYPE: PRT  
ORGANISM: NO. US20030233675A1loc punctiforme  
US-10-369-493-19944  
Query Match 30.1%; Score 46; DB 12; Length 584;  
Best Local Similarity 42.9%; Pred. No. 1.6e+02;  
Matches 9; Conservative 1; Mismatches 11; Indels 0; Gaps 0;  
CY 5 GNVIGNRHHYDGKXFYTKQPN 25  
DB 415 GRVVSANFSYDGKQIITGSGN 435  
RESULT 11  
US-09-764-891-4533  
Sequence 4533, Application US/09764891  
Publication No. US20030077808A1

GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC006  
CURRENT APPLICATION NUMBER: US/09/764,891  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PAM or file wrapper  
NUMBER OF SEQ ID NOS: 10231  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 4533  
LENGTH: 147  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-764-891-4533  
Query Match 29.7%; Score 45.5; DB 11; Length 147;  
Best Local Similarity 44.4%; Pred. No. 45;  
Matches 12; Conservative 6; Mismatches 4; Indels 5; Gaps 3;  
CY 6 NYIQNRHY-DG-KRFTYKQFN--KP 27  
DB 61 HYDEKNWYCDGYDRRFYETIINGLKP 87  
RESULT 12  
US-10-264-049-3481  
Sequence 3481, Application US/10264049  
Publication No. US2004005579A1  
GENERAL INFORMATION:  
APPLICANT: Bize et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PA133PI  
CURRENT APPLICATION NUMBER: US/10/264,049  
CURRENT FILING DATE: 2002-10-04  
PRIOR APPLICATION NUMBER: PCT/US01/18569  
PRIOR FILING DATE: 2001-06-07  
PRIOR APPLICATION NUMBER: US 60/209,467  
PRIOR FILING DATE: 2000-06-07  
NUMBER OF SEQ ID NOS: 4360  
SOFTWARE: Patentin Ver. 3.1  
SEQ ID NO 3481  
LENGTH: 163  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (136)  
OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (138)  
OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (141)  
OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (148)  
OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids  
US-10-264-049-3481  
Query Match 29.4%; Score 45; DB 12; Length 163;  
Best Local Similarity 60.0%; Pred. No. 59;  
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
CY 13 HYDGKXFYTKQFNKP 27  
DB 146 HXGKWCILTYFNKP 160  
RESULT 13  
US-10-264-237-1704

```

; Sequence 1704, Application US/10264237
; Publication No. US2004009491A1
; GENERAL INFORMATION:
; APPLICANT: Bize et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P1331P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 1704
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (81)-
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (178)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (195)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (241)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (291)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-237-1704

Query Match          29.4%; Score 45; DB 12; Length 308;
Best Local Similarity 56.2%; Pred. No. 1.2e+02;
Matches 9; Conservative 2; Mismatches 1; Indels 4; Gaps 1;

QY 5 GNYIQRN---FHYDG 16
DB 109 GNYIELNENDPHYNG 124

RESULT 14
US-09-867-550-1188
; Sequence 1188, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Foad
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and
; OF INVENTION: Thereby
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1188
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-867-550-1188

```

```

Query Match          29.4%; Score 45; DB 9; Length 318;
Best Local Similarity 56.2%; Pred. No. 1.2e+02;
Matches 9; Conservative 2; Mismatches 1; Indels 4; Gaps 1;

QY 5 GNYIQRN---FHYDG 16
DB 183 GNYIELNENDPHYNG 198

RESULT 15
US-09-764-864-842
; Sequence 842, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 842
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-842

Query Match          29.4%; Score 45; DB 10; Length 422;
Best Local Similarity 56.2%; Pred. No. 1.6e+02;
Matches 9; Conservative 2; Mismatches 1; Indels 4; Gaps 1;

QY 5 GNYIQRN---FHYDG 16
DB 159 GNYIELNENDPHYNG 174

Search completed: February 10, 2004, 17:01:31
Job time : 35 secs

```